SportRχiv

Part of the Society for Transparency, Openness and Replication in Kinesiology (STORK)

Preprint
not peer reviewed

Acute to random chronic workload ratio is 'as' associated with injury as acute to actual chronic workload ratio: time to dismiss ACWR and its components

Received: March 3, 2020
10.31236/osf.io/e8kt4
For correspondence:
Franco.Impellizzeri@uts.edu.au

Impellizzeri FM,[1]

Woodcock S,[2]

Coutts AJ,[1,3]

Fanchini M,[4]

McCall A,[5]

Vigotsky AD,[6]

1, University of Technology Sydney (UTS), Human Performance Research Centre, Faculty of Health, Sydney, NSW, Australia

2, Carlton Football Club, Carlton North, Victoria, Australia

3, University of Technology Sydney (UTS), School of Mathematical and Physical Sciences, NSW, Australia

4, AS Roma Performance Department, AS Roma Football Club, Roma, Italy

5, Arsenal Performance and Research Team, Arsenal Football Club, London, UK

6, Departments of Biomedical Engineering and Statistics, Northwestern University, Evanston, IL

Please cite as: Impellizzeri FM, Woodcock S, Coutts AJ, Fanchini M, McCall A, Vigotsky AD, (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. *SportRχiv*.
https://doi.org/10.31236/osf.io/e8kt4

All authors have read and approved this version of the manuscript. This article was last modified on March, 2020.

Authors *@francoimpell, @avigotsky, @AaronCoutts, @Maurizi13707619* and *@Alan_McCall_* can be reached on Twitter.

ABSTRACT

Aim: The aim of this study is to assess the presence and implications of statistical artefacts created by a commonly used indicator of injury risk in both research and practice: the ratio between acute workload (AL) and chronic workload (CL), named ACWR.

Methods: Using previously published data, we generated a contrived ACWR by dividing the AL by fixed and randomly generated CLs, and we compared these results to real data. We also reproduced previously reported subgroup analyses, including dichotomising players' data above and below the median CL. Our analyses follow the same, previously published modelling approach.

Results: After reproducing the original analyses with only the ACWR showing effects compatible with higher injury risk (odd ratios, OR: 2.45, 95%CI 1.28 to 4.71), we demonstrated similar findings by dividing AL by the "contrived" fixed and randomly generated CLs: OR=1.95 (1.18 to 3.52) dividing by 1510 (average CL); and OR using random CL= 1.53 (mean), ranging from 1.16 to 2.07. Random ACWR calculated reducing the variance of the original AL further inflated the ORs (mean OR=1.89, from 1.42 to 2.70). ACWR causes artificial reclassification of players compared to AL alone. Finally, neither ACWR nor AL alone confer a meaningful predictive advantage to an intercept-only model, even within the training sample (*c*-statistic = 0.574/0.544 vs. 0.5 in both ACWR/AL and intercept-only models, respectively).

Discussion: ACWR is a rescaling of the explanatory variable (AL, numerator), in turn magnifying its effect estimates and decreasing its variance despite conferring no predictive advantage. Other ratio-related transformations (e.g., reducing the variance of the explanatory variable and unjustified reclassifications) further inflate the OR of AL alone with injury risk. These results also disprove the etiological theory behind this ratio and its components. We suggest ACWR be dismissed as a framework and model, and in line with this, injury frameworks, recommendations, and consensus be updated to reflect the lack of predictive value of and statistical artefacts inherent in ACWR models.

Keywords: training load; injuries; ratio; artefacts; acute; chronic.

INTRODUCTION

The number of studies examining the relation between training load and injuries in athletic populations have grown exponentially in recent years, and at present, there are over 100 studies on the topic.[1-4] To find an association between training load and injuries, various measures of training exposure have been created. The most popular metric, commonly used as a gold standard reference "model" for several international guidelines, is the acute:chronic workload ratio (ACWR).[1,5-9] This ratio is obtained by dividing a 'fatigue' component by a 'fitness' component. The 'fatigue' component is represented by the acute workload (AL), commonly calculated using the workload of the week preceding the injury, while the 'fitness' component is represented by the chronic workload (CL), which is the average workload of the four weeks preceding the injury.[6,7] The AL compared to the CL as measured using this ratio is widely considered to reflect the risk of injury in athletic populations.

ACWR has recently taken sports science and medicine by storm. It has consistently been claimed that the ACWR is associated with injury risk,[1,5,7,10-13] making it a useful metric to reduce the injury risk or prevent injury.[7] This metric has been popularised by several editorials and consensus in high impact factor sport science and medicine journals.[5,7,10-13] Speaking to their influence, these papers are amongst the most highly cited in the field. The rise in the attention received by "load management" in professional practice has also been fuelled by these studies. The influence of ACWR has even bled into the international circuit; it is being used in the development international guidelines and consensus statements by leading organisations such as the International Olympic Committee (IOC).[12] ACWR is ubiquitous, and is included in national athlete management systems and commercially available software under the assumption that it can help reduce injuries.

Adaptations of ACWR have been proposed using different ways to calculate the AL and CL, such as the exponentially weighted moving average (EWMA),[14,15] coupled or uncoupled (AL included or not in the CL calculation),[16] and different time windows.[17,18] Regardless of the method, all have been suggested to work (i.e., are associated with injury); yet, all have conserved a common characteristic: they are all ratios. Researchers have warned about the use of the ACWR because of a ratio's failure to normalise the

numerator by the denominator and the risk of artefacts (i.e. it adds unnecessary noise).[19] However, not only did these warnings not gain traction,[2] they have been largely ignored, and in doing so, ignore issues that have been highlighted by statisticians for decades.[19-21]

The aim of this study is to explicate the ratio effects of the ACWR. By using a previously published dataset from professional football players, where originally, a relation between ACWR and injury was reported, we demonstrate the artefacts introduced through the use of a ratio.

METHOD

Dataset

Although this demonstration could also be achieved with simulated data, we used a previously published dataset to show the impact on real world data and results. The details of the data collection can be found in the original manuscript that has been made freely accessible online by the publisher. The manuscript does not comply with the Strengthening the Reporting of Observational Studies in Epidemiology or Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis recommendations for reporting since it is a methodological study.[22,23]

Participants

Briefly, the players' individual training load was collected on a professional Italian Serie A team, on 34 players (age: 26 ±5 y; height: 182 ± 5 cm; body mass: 78 ± 4 kg) over 3 competitive seasons (2013/14, 2014/15, and 2015/16). The dataset was the same used by Fanchini et al.,[24] but we deleted the individual player loads with missing data to allow better comparisons between analyses and to avoid any missing data imputation potential influence. 36 weekly loads were excluded out of 1955 (1.8%) and two injuries out of 72. The final dataset included 1919 individual weekly sessions and 70 injuries. Descriptive data are presented in Table 1.

Training load and injury

Internal training load was quantified using the session Rating of Perceived Exertion (RPE) method; that is, by multiplying the training session duration by the corresponding RPE value determined using the Borg's CR10 scale.[25] Using these training loads, we calculated:

1. AL, average training load of 1 week preceding the injury;
2. CL, rolling averages of 4, 3 and 2 weeks preceding the injury including the AL in the calculation of CL (i.e., coupled) as in the original study;
3. CL, rolling averages of preceding the injury without including the AL in the calculation of CL (i.e., uncoupled) for the calculation of #7:
4. ACWR, ratio between AL and CL;
5. Contrived ACWR, ratios between AL and fixed and randomly generated values of CL;
6. Week-to-week change, difference between ALs the two weeks preceding the injury;
7. AL-CL difference, absolute AL-CL difference (coupled and uncoupled);

Data were also categorised using quartiles, and two groups based on the median CL value were also determined. Injuries were classified according to international guidelines,[26] and recorded by medical staff. Only non-contact, time loss injuries were used for the analysis.

ACWR variations

For this study we created contrived ACWRs using fixed CLs: 500, 1000, 1510 (corresponding to the average CL of the sample), 2000, and 2500. These represent the effect of a simple linear rescaling of the AL, as no variance is contributed by the CLs.

In addition to fixed CLs, we also calculated ACWR values using independently and identically distributed randomly generated data (from a normal distribution). We first generated samples having the same mean and standard deviation (SD) of the original sample, and two sets with lower (SD original 282/2=141 AU) and higher (SD original 282+141=423 AU) values than the original SD. We generated 25, 20 and 20 ACWR values for each condition, respectively. Second, we performed these simulations (100/condition) for a range of mean chronic workloads (500 through 2500, step size = 100; original mean = 1510) and coefficients of variation (CV) (5% through 50%, step size = 5%; original CV ≈ 20%). In doing so, we covered a large sampling space and investigated the effects of different magnitudes and spreads of CLs, which were independent of time, individual, and thus true CL. Estimates from these random models were calculated using trimmed mean, excluding the top and bottom 10% of simulated estimates; otherwise, there were instances of massive outliers which shifted the mean by greater than one order of magnitude.

Statistical analysis

All analyses were consistent with that of the previously published study: generalised estimating equations (GEE) with a logistic link function, robust variance estimation, and an exchangeable working correlation matrix. We note that GEEs were not chosen because we considered them the best way to analyse these kinds of observational studies, but rather, to illustrate the potential for artefacts in a way that is congruent with the analytical approaches present in the

literature. In addition to assessing the resulting odds ratios (OR), we also assessed proper scoring rules (Brier), *c*-statistics (equivalent to area under the receiving operator characteristic curve), and the estimated probabilities of injury. If the parameter estimate is statistically significant but the model itself does not fit the data well, the overall value of the parameter is unclear. We contend that, ultimately, we are interested in modelling injury risk, and as such, the model should fit the outcome well, statistically significant parameter or otherwise. Given the low sample size,[27] the absolute values of the Brier scores were not interpreted; instead, the Brier scores were calculated for comparison purposes. Finally, although not shown, similar results were obtained with other traditionally used analyses and variations (e.g. GEE using Poisson and changing working correlation matrix, or logistic regression without accounting for repeated measures, etc.).

Mean difference and 95% confidence intervals (95%CI) were also calculated for comparing injured and non-injured players.

RESULTS

Descriptive data of the explanatory variables used in this study are presented in Table 1, including the quartiles used for categorising. Depictions of injuries as a function of AL, CL, and ACWRs are presented in Figure 1.

The results of the GEE using the original data but without using the ACWR, are presented in Table 2. Importantly, the results of the original model (ACWR, 4 weeks) indicate ACWR as a predictor confers no predictive advantage to an intercept-only model, even within the training sample (Brier score = 0.035 vs. 0.035; *c*-statistic = 0.574 vs. 0.5 in ACWR and intercept-only models, respectively; AL alone was identical to ACWR). Despite this, we investigated and quantified the role of different workloads in other models.

Some associations are what would classically be considered "statistically significant" ($p < 0.05$). However, the odd ratios (OR) were negligible *or* their Brier scores and *c*-statistics were comparable to an intercept-only model. These results are similar to the ones of the original publication[24] and directly follow from the distributions of the raw data, which indicate that injuries are relatively evenly dispersed across AL, CL, and ACWR (Figure 1). The results using the original ACWR, the ACWR created using fixed values of CL, and dichotomising the players' data in high and low CL are presented in Table 3. All the ORs from the association between injury and ACWR values were in the direction of increased injury risk, with the exclusion of the analysis of the high CL group.

Average point estimates (ORs) obtained by generating random CL for the calculation of the ACWRs are presented in Figure 2 (large space) and Appendix 1 (small space, relative to original sample). The direction of association was generally consistent across random models, but the magnitude was a function of the mean CL and the coefficient of variation of the CL. In all cases, the models had poor predictive performance, much like the original model. The ORs obtained from GEE using ACWR calculated from random CL with the same SD (282 AU) ranged from 1.16 to 2.07. Using half the original SD (141 AU) the ORs ranged from 1.41 to 2.70. Increasing the SD to 423 AU, the ORs ranged from 0.89 to 1.31. Details (p values and CIs) of this analysis are presented in Appendix 1.

In Table 4 we presented the comparison between injured and uninjured players' data for AL, AL divided by the a fixed value corresponding to the average original CL (1510 AU), ACWR from 4 to 2 weeks and AL or ACWR for the two groups classified according to the median CL values. Differences between groups are presented with the corresponding 95%CI.

Crosstab showing the classification of the players' data point according to four categories of AL and four categories of ACWR are presented in the Table 5. Crosstabs are for the two groups based on median CL separately. Number of injuries for each ACWR category are also presented (for low CL group we also indicated the original AL category). Categories have been created as quartiles (values presented in Table 1). The within player relation between AL and CL is presented in Appendix 2.

DISCUSSION

We systematically evaluated the ACWR concept by comparing it to an acute-to-random workload. When used in training load–injury models, the ACWRs creates remarkable statistical artefacts in the effect estimates. Here, we focus on the outcomes generated by these artefacts and provide some preliminary explanations. These findings demonstrate that when ACWR is used as an explanatory variable, results are always influenced by artefacts and artificial alterations. We have also shown that, depending on the characteristics of the sample (injury and data distribution), these artefacts can result in associations that can be statistically significant or compatible with increased or decreased injury risk.

The theory behind the use of the ACWR states that, when the AL exceeds the CL, an athlete is underprepared and hence at higher injury risk. The ACWR would indicate "both the athlete's risk of injury and preparedness to perform".[7] This concept was linked to the Banister model, which used

two components: fitness (represented by the chronic load) and fatigue (represented by acute workload). The ACWR has also been linked to another similar metric, Total Stress Balance, also calculated using the fitness and fatigue components of Banister.[5-7] [7] However, while these two reference models were additive, for reasons unbeknownst to the authors, ACWR relies on a ratio. Moreover, it was suggested that the negative effect of increasing load is greater when the CL is lower.[28] These models are conceptually different insofar as the Banister model investigates the effect of fatigue while controlling for fitness, while ACWR implies the absolute effect of fatigue changes with fitness.

The ACWR approach can be reframed similarly to the Banister model by way of stratifying observations based on CL. We tested the utility of these stratification procedures by reanalysing a previously published dataset. In doing so, we did not find any meaningful associations (Table 2). We also examined the independent effects of AL, CL, and their interaction; again, we did not find any meaningful associations with injury risk, suggesting that controlling for CL does not confer a meaningful advantage. These results, in addition to those of the original study, apparently supported this association since ACWR was the only variable found significantly related to injury risk (Table 3). Stratifying or controlling by CL does not seem to be advantageous, which suggests one of two conclusions: (1) ACWR appropriately captures the construct we are attempting to model (injury risk), or (2) CL does not contain any useful information. We performed further analyses to test these competing explanations.

If the proposed etiological theory of ACWR was correct, then dividing the individual AL by a contrived CL (i.e., a value not corresponding to the real CL of each player) should produce disparate results from ACWR, since it violates the underlying etiological theory. Therefore, we started by simply dividing the AL of all the players by the same value (i.e., the average CL value, 1510 AU), and this 'contrived' CL replaced the players 'real' CL. Rather than an ACWR, this is an "acute to fixed workload ratio." Surprisingly, the OR was 1.95 (1.08 to 3.52), which is just slightly lower than the OR from the ACWR model (2.45, 1.28 to 4.71). Importantly, our analysis still suggested that the acute:'fixed' workload ratio performed similarly and still yielded a "statistically significant" association with injury risk. We repeated the analysis with other 'contrived' fixed values, and intuitively, by increasing or decreasing the denominator, the *p*-values remained the same, while the estimates increased or decreased (see Table 3).

Therefore, we generated random CL samples with a similar mean and SD of the original data, which is the equivalent of dividing the AL of a player by the CL of another hypothetical random teammate. Since this has no logical basis, it can be conceived as a null model to assess the value of CL. Once again, we found associations between these contrived ACWR values and injury. From these data, we could call findings based on ACWR into question—the ACWR appears to simply be a linear rescaling of AL alone and provides no additional information. What is more, this finding calls into question theory behind the ACWR, which may have arisen as a *post hoc* theory from a statistically significant predictor rather than one borne and hypothesized *a priori* from a deep theoretical framework (i.e. HARK-ing, Hypothesised After Results are Known). Undeniably, the results strongly demonstrate that CL does not reflect "preparation' of the players and confer no added value, as even randomly-generated, ACWRs with contrived CLs perform similarly to ACWRs with true CLs.

But why does this happen? Actually, the answer is quite simple. By dividing the numerator (AL) by a number, the researchers have just rescaled the numerator. The parameter estimates from the model correspond to a one unit increase in the explanatory variable. When rescaling, the unit is still one, but it now corresponds to a different quantity in the explanatory variable. The new unit of the ACWR indeed corresponds to the amount of the CL; i.e., 1 unit = 1 CL. If the CL is on average 2000 (AU, meters, etc.), the new estimate is now 2000 times the estimate corresponding to 1 in the original scale (e.g., 1 AU or 1 m). In other words, the scale of the parameter estimate must offset the rescaling of the numerator. Since measures like ORs (or relative risk, etc.) are multiplicative, the new effect is even greater. That is, the model estimates log(OR) as the parameter, which is exponentiated to obtain the OR. What is multiplicative on the log(OR) scale is exponential when brought back to the original OR scale, and thus, the OR is raised to 2000. Whatever the number γ of the denominator, the "new" OR will be the one of the numerator raised to γ. To draw a concrete example, if we have OR=1.001 for 1 m, but we want to refer the OR for 1 km, we can divide the original variable by 1000. The new OR will be: 1.001^{1000}. Simply, this transformation follows from the laws of logarithms and magnifies the magnitude of the OR estimated using AL alone; when predictor units change, parameter estimate units change accordingly.

The rescaling of AL can be, and has been, used to produce more impressive parameter estimates. Through simple transformations, a difference in AL will generate impressive effects when using the ratio. Indeed, in the sub-analysis performed to reflect previous studies (e.g., dichotomising player data based on a median split of CLs), we found appreciable differences in the AL between injured versus uninjured player data. As shown in Table 4, the injured players in the low CL group have greater ALs. As for the whole sample, there is a negligible effect of AL, even if

statistically significant (ORs from 1.000 to 1.001). However, when the ACWR was used, the OR increases exponentially to 2.9 (1.6 to 5.1, Table 4). As further confirmation, an even greater OR was obtained dividing by the AL by 1510 AU compared to "real" CL (3.5, 1.7 to 7.3). Once again, the underlying etiological theory (chronic load "protective") has nothing to do with the reasons for these results—rather, these results follow directly from the mathematics underlying the statistical model.

Dividing the AL by the CL not only changes the properties of the mean, but also the variance. Because CL is a temporally smoothed version of the AL, it has lower variance, and thus, when using it as the divisor, it creates a variable with a lower coefficient of variation and smaller mean than AL alone. This results in a greater parameter estimate, and also influences the *p*-values and CIs. By generating random CLs with a mean similar to the original sample, but enlarging or restricting the SD, the point estimates, CIs and *p*-value are changed compared to the AL alone (Appendix 1). Specifically, when the SD of the randomly generated CL data was lowered, the *p*-values decreased and ORs increased. This can be also seen in the Figure 2B that shows the ORs generated by with different CL means and coefficients of variation (i.e. SD). While these results can be obtained using both coupled and uncoupled ACWRs, the coupled ACWR has additional issues. Since the numerator is included in the denominator, the variance of the ratio will inevitably be smaller. This additional artefact, caused by shrinking the SD, also explains why the use of the CL calculated using the average of more weeks (or days) is exploits this artefact. Using a rolling average in the denominator creates a positive correlation between the numerator and denominator. The result of this is that large values of AL are attenuated by division by larger CLs, hence reducing the variability of the ratio.[29]

General Problems with Ratios in Predictive Models

While the aforementioned consequences of the ratio transformation are sufficient to invalidate the ACWR and the etiological theory behind it, we highlight a further problem generated by transforming data into a ratio as it results in a reclassification. First, we note the differences in properties between multiplicative (ratio) and linear scales. Indeed, Curran-Everett and others[20,30] have warned against the use of ratios and percentages in such analyses, in part because the values depend on the direction of the comparison. For example, if training load is reduced from 1000 to 800 (meters, AU, etc.), the relative decrease will be 20%, while if you increase from 800 to 1000, the relative increase will be 25%. These multiplicative changes are in contrast to additive ones, which are linear.

Second, because ACWRs are a proportion and thus sensitive to the denominator, individual players with low absolute ALs tend to have greater ACWRs, resulting in model miscalibration. For example, injured players with the lowest AL values tend to move in the higher category of the ACWR. This is evidenced by Table 5, where the data of high and low CL groups are presented separately to reproduce a typical dichotomisation of the data used in previous studies. Individual data with the lowest levels of AL belonging to the first quartile (< 1261 AU) moved into the higher ACWR categories (226 individual data, 57%); similar reclassification can be seen in the other categories. This shift was more prevalent in the low CL group since dichotomising by CL means also separating by AL. Lower AL values are more likely to produce greater ACWR values when AL increases, since it represents a larger proportion of the denominator (CL). Indeed, there is an obvious relation between AL and CL (Figure 2). Similar subgroup analyses have been used to support the claim that high CL is protective while low CL predisposes athletes to injuries when "spikes" of workload occur: studies have reported a stronger association between ACWR and injury risk at low compared to high CL.[28,31] Performing the same analysis in this sample (n.b. this was not done in the original publication), the ACWR was also found to be associated with greater injury risk for the low CL group only, thus seemingly supporting previous findings. While one may think that this reclassification is appropriate, since it appears to account for the increase "impact" of increasing load when the player is not "prepared" (i.e. low CL), we have already shown that this theory (protection or predisposition) does not stand since the CL itself magnifies and smooths the effect of AL effect estimates (i.e. just a rescaling number). If this theory held true (low CL predisposing), we would have found an association between AL or AL-CL change and the levels of CL (Table 2), and we would not have found similar results when using the contrived CL values. This did not occur. Rather, we also observed that each of the 12 injured players in the high ACWR group came from lower categories of the AL. Three were from the first quartile, one from the second and eight from the third. This example showed that the reclassification is artificial and the ratio gives more "weight" to absolute changes at low workloads and more "weight" when the workloads increase rather than when they decrease. This explains why this reclassification occurred more in the low than high CL group. Moreover, from a statistical theory standpoint, the "split"-based analysis implies an interaction between ACWR and CL; however, not only does this simplify to AL alone, but the OR=1.0 (1.0, 1.0) for their interaction. Although this adjustment does not have as large an impact as rescaling, it still biases the results and creates artificial differences between injured and uninjured. Reclassification of 12 injured players out of 36 in the high ACWR has a clear effect on the results and, as in previous studies, also on the calculations of other figures such as injury rate. While in the past, similar results have been used to support the predisposing effect of low CL. However, the evidence and

logic we present suggest this is, instead, another result of the combination of statistical artefacts and noise added by the ratio, also causing re(mis)classification.

The ACWR creates artefacts generated by the combination of the aforementioned factors altering and magnifying the effects of the AL (numerator). Depending on the relation between AL and injuries, the effect estimates are increased and, depending on the distribution of the denominator, they are further inflated. Therefore, the ACWR values calculated from different smoothing averages (e.g., 2 to 4 weeks) with the highest value and lowest SD (Figure 2B) will magnify the estimates and influence the *p*-values and CIs. The use of a ratio and further reducing the variance of the explanatory variable using other smoothing strategies—such as the EWMA, as suggested and used in some studies—suffers from the same problems. In addition, they also are not conceptually superior since the starting idea of a CL-AL interaction is not supported, but rather is just an artefact (whatever the mathematical "strategy" to calculate the "fatigue" and "fitness" components). Studies showing the superiority of ACWR based on EMWA, or the "equivalence" between coupled and uncoupled, confirm that these methods produce the same artefacts.[14,16,32] Similarly, explorative studies trying to find the best combination of AL and CL time windows to "optimize" parameter estimates may just be optimizing these artefacts (involuntary p-harking).[17] Hold-out samples should be used to evaluate the effects of optimization, and prediction/model fit should be assessed rather than parameter estimation alone. Similarly, most arbitrary pre analytical data "treatment" also amplifies these artefacts by, for example, changing the variance of the AL, CL, and their ratio (e.g. deleting CL below 1 or 2 SD, single imputation, etc.).[8,17,28]

It may seem from the arguments we put forth that the "key" metric to focus on is the acute workload. Although we will not address this topic in detail here, it is not so straightforward. Simply comparing the AL (or any other potential factor) of injured versus non-injured is not sufficient as the studies from which these data come are prone to several potential biases well known in epidemiology.[33,34] Therefore, it is not a question of "statistical analysis" or creating new metrics calculated from each other, but rather design and conceptually selecting explanatory variables based on a proper conceptual and theoretically sound framework, all while controlling for confounding factors. Moreover, it is essential that the predictive performance of these models be assessed out-of-sample. If these models are predictive of injury in hold-out samples, experimental approaches to manipulating the predictors (e.g., acute load) should be employed to assess the causal nature of the relationship. This approach is essential for causal inference, which is arguably the tacit aim of these studies. Indeed, as a evidence of this causal interpretation, other than the overinterpretation of the studies themselves, we now have international guidelines and consensus suggesting how to manipulate these prognostic factors (training load metrics) to reduce the injury risk, which assumes a causal effect (i.e., a perturbation in *x* results in a change in *y*). Importantly, this assumption has been made in the complete absence of any attempts to estimate causal effects and based on results determined by artefacts due to data transformations. The interpretation should always be based on and commensurate with the real nature and goal of the study (descriptive, predictive, causal).

Predictive Value of ACWR and Acute Workload

Although not the primary purpose of this work, we briefly explored the in-sample predictive value of the ACWR. Despite having a statistically significant and large OR, ACWR confers no predictive advantage with respect to injury risk. Proper scoring rules are virtually identical between ACWR and an intercept-only model (both Brier scores = 0.0351), and the intercept-only model has a slightly greater *c*-statistic than the ACWR model (0.574 vs. 0.5). In the ACWR model, the average probability of injury of those who were injured was 0.039 (Figure 4). We replicated the aforementioned analyses using AL-only, and the results were identical, with the exception of the *c*-statistic, which went from 0.574 to 0.544. From a predictive standpoint, when used in isolation, neither AL nor ACWR contain useful information, even when assessed in the training sample.

Conclusion

We are confident that most of these errors that have been made in previous studies were unintentional. It is also reasonable that the authors believe that the reported relation between training and injury was authentic, and that the etiological theory created to support the ACWR and its components was rational. However, as the ACWR model fitted popular beliefs so well, it became a self-fulfilling prophecy and lowered scientists willingness to critically evaluate the construct. The selection of candidate prognostic factors may benefit from explorative studies, but we urge scientists to avoid procedures that may produce statistical artefacts and that focus on the dichotomization of effects (e.g., null hypothesis significance testing). In the current study, we have demonstrated using published data and simulations that:

- the etiological theory developed to explain the relation found in some studies between ACWR and injury risk is not supported;
- the ratio is a rescaling procedure, exponentially magnifying the effect of the AL;
- a ratio using averages of the numerator as the denominator will have a lower SD, such that a one unit increase in the new explanatory variable will correspond to a higher ORs;

- the ratio also causes artificial and non-physiologically justified reclassifications, further influencing the results;
- neither ACWR nor AL contain useful information for predicting injury;
- the findings based on ACWR reported in the literature are therefore all affected by artefacts that, depending on the data characteristics, resulted in negative, positive, or no associations (in this dataset positive associations).

Practical applications

The ACWR and its components should be dismissed. Moving forward, time should be focused on selecting and identifying appropriate proxy measures and developing reasonable causal assumptions. Creating new metrics without conceptual reference models and relying on statistical significance, especially for prediction, should be avoided. The results of previous studies should be reconsidered, and authors and editors should make efforts to correct the erroneous messages that were disseminated, and their associated theoretical frameworks should be revised. Finally, international and national organizations and athlete management system that base their recommendations on the results of these studies should revise their recommendations, acknowledging these artefacts and lack of predictability.

ACKNOWLEDGMENTS

The Authors would like to thank Lorenzo Lolli for the suggestions, Marco Riggio and Ermanno Rampinini for the support in the data extraction, preparation and data sharing.

REFERENCES

1. Maupin D, Schram B, Canetti E, Orr R. The Relationship Between Acute: Chronic Workload Ratios and Injury Risk in Sports: A Systematic Review. *Open Access J Sports Med.* 2020;11:51-75.
2. Griffin A, Kenny IC, Comyns TM, Lyons MJSM. The Association Between the Acute:Chronic Workload Ratio and Injury and its Application in Team Sports: A Systematic Review. *Sports medicine (Auckland, NZ).* 2019.
3. Drew MK, Blanch P, Purdam C, Gabbett TJ. 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. *Br J Sports Med.* 2017;51(7):618-619.
4. Eckard TG, Padua DA, Hearn DW, Pexa BS, Frank BS. The Relationship Between Training Load and Injury in Athletes: A Systematic Review. *Sports medicine (Auckland, NZ).* 2018;48(8):1929-1961.
5. 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(8):471-475.
6. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med.* 2016;50(5):273-280.
7. Gabbett TJ, Hulin BT, Blanch P, Whiteley R. High training workloads alone do not cause sports injuries: how you get there is the real issue. *Br J Sports Med.* 2016;50(8):444-445.
8. Hulin BT, Gabbett TJ, Lawson DW, Caputi P, Sampson JA. 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):231-236.
9. Hulin BT, Gabbett TJ, Blanch P, Chapman P, Bailey D, Orchard JW. Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *Br J Sports Med.* 2014;48(8):708-712.
10. Gabbett TJ, Whiteley R. Two Training-Load Paradoxes: Can We Work Harder and Smarter, Can Physical Preparation and Medical Be Teammates? *Int J Sports Physiol Perform.* 2017;12(Suppl 2):S250-s254.
11. Bourdon PC, Cardinale M, Murray A, et al. Monitoring Athlete Training Loads: Consensus Statement. *Int J Sports Physiol Perform.* 2017;12(Suppl 2):S2161-S2170.
12. Soligard T, Schwellnus M, Alonso JM, 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(17):1030-1041.
13. Pluim BM, Drew MK. It's not the destination, it's the 'road to load' that matters: a tennis injury prevention perspective. *Br J Sports Med.* 2016;50(11):641-642.
14. Murray NB, Gabbett TJ, Townshend AD, Blanch P. 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):749-754.
15. Williams S, West S, Cross MJ, Stokes KA. Better way to determine the acute:chronic workload ratio? *Br J Sports Med.* 2017;51(3):209-210.
16. Gabbett TJ, Hulin B, Blanch P, Chapman P, Bailey D. To Couple or not to Couple? For Acute:Chronic Workload Ratios and Injury Risk, Does it Really Matter? *Int J Sports Med.* 2019;40(9):597-600.
17. Carey DL, Crow J, Ong KL, et al. Optimizing Preseason Training Loads in Australian Football. *Int J Sports Physiol Perform.* 2018;13(2):194-199.
18. Malone S, Owen A, Mendes B, Hughes B, Collins K, Gabbett TJ. 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.* 2018;21(3):257-262.
19. Lolli L, Batterham AM, Hawkins R, et al. The acute-to-chronic workload ratio: an inaccurate scaling index for an unnecessary normalisation process? *Br J Sports Med.* 2018.
20. Curran-Everett D. Explorations in statistics: the analysis of ratios and normalized data. *Adv Physiol Educ.* 2013;37(3):213-219.
21. Kronmal RA. Spurious Correlation and the Fallacy of the Ratio Standard Revisited. *Journal of the Royal Statistical Society.* 1993;156(3):379-392.
22. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med.* 2007;147(8):W163-194.
23. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.
24. Fanchini M, Rampinini E, Riggio M, Coutts AJ, Pecci C, McCall A. Despite association, the acute:chronic work load ratio does not predict non-contact injury in elite footballers. *Sci Med Football.* 2018;2(2):108-114.

25. Impellizzeri FM, Rampinini E, Coutts AJ, Sassi A, Marcora SM. Use of RPE-based training load in soccer. *Medicine and science in sports and exercise.* 2004;36(6):1042-1047.
26. 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.
27. Wilks DS. Sampling distributions of the Brier score and Brier skill score under serial dependence. *QJR Meteorol Soc* 2010;136(653):2109-2118.
28. Bowen L, Gross AS, Gimpel M, Bruce-Low S, Li FX. 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. *Br J Sports Med.* 2019.
29. Frishman F. On the Arithmetic Means and Variances of Products and Ratios of Random Variables. In: Patil GP, Kotz S, Ord JK, eds. *A Modern Course on Statistical Distributions in Scientific Work.* . Vol 17. Dordrecht: Springer; 1975.
30. Curran-Everett D, Williams CL. Explorations in statistics: the analysis of change. *Adv Physiol Educ.* 2015;39(2):49-54.
31. Colby MJ, Dawson B, Peeling P, et al. 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.* 2017;20(12):1068-1074.
32. Coyne JOC, Nimphius S, Newton RU, Haff GG. Does Mathematical Coupling Matter to the Acute to Chronic Workload Ratio? A Case Study From Elite Sport. *Int J Sports Physiol Perform.* 2019:1-8.
33. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health.* 2004;58(8):635-641.
34. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2012.

Table 1. Descriptive data of the explanatory variables used in the analyses

	Mean	Median	SD	Range	Minimum	Maximum	Percentiles		
							25	50	75
Acute Load	1526	1542	442	2977	120	3097	1261	1542	1851
Chronic Load (4 weeks)	1510	1523	282	1820	437	2257	1328	1523	1718
Chronic Load (3 weeks)	1511	1524	299	1843	518	2362	1304	1524	1727
Chronic Load (2 weeks)	1516	1532	333	2366	280	2646	1287	1532	1752
Week to week difference	20	1	620	4187	-1990	2197	-355	1	377
Acute - Chronic load (coupled)	15	26	362	2566	-1257	1309	-203	26	249
Acute - Chronic load (uncoupled)	22	42	457	3405	-1594	1811	-246	42	315
ACWR 4 weeks	1.016	1.017	0.270	2.841	0.129	2.970	0.867	1.017	1.166
ACWR 3 weeks	1.014	1.012	0.257	2.473	0.140	2.612	0.871	1.012	1.151
ACWR 2 weeks	1.010	1.000	0.243	1.842	0.158	2.000	0.886	1.000	1.129

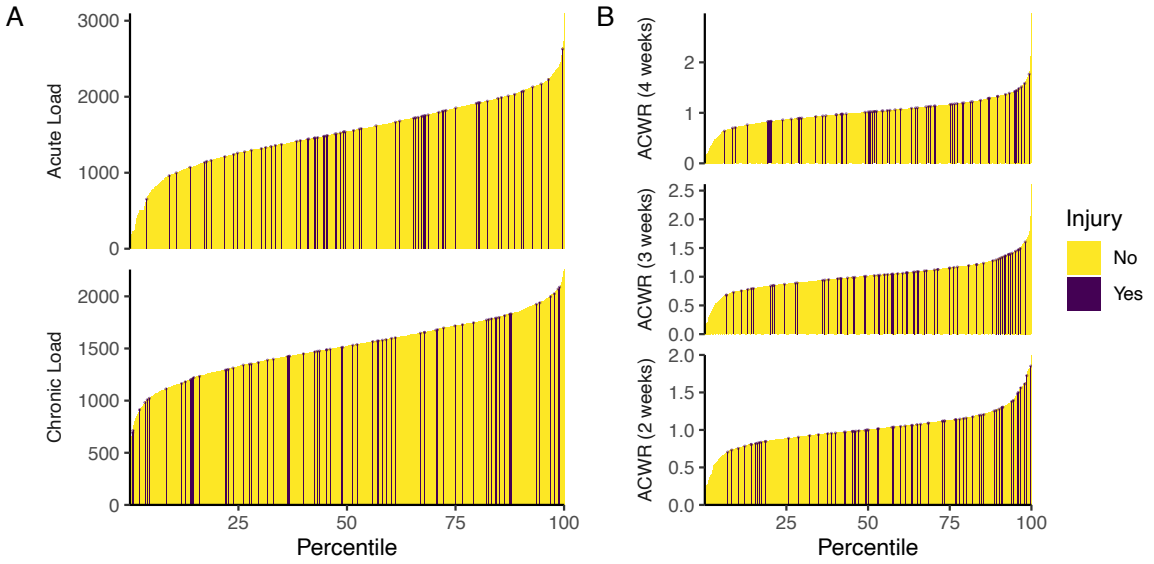

Figure 1. Distributions of loads and load ratios for data points where injuries did and did not occur. (**A**) depicts the distribution of acute (top) and chronic (bottom) loads for all data points included. (**B**) illustrates the distributions of acute-to-chronic workload ratios for different lengths of time. The yellow indicates data points where no injury occurred, while dark purple indicates data points where injuries did occur. Note, the injured data points are relatively uniformly dispersed, indicating that neither the raw loads or load ratios will be predictive of injury risk.

Table 2. Parameters of various models estimated from original data.

	B	Std. Error	Wald	Exp(B)	95% Confidence Intervals	
			Sig.	OR	Lower	Upper
Acute workload (AL)	0.000	0.000	0.027	1.000	1.000	1.001
Week to week AL difference	0.000	0.000	0.022	1.000	1.000	1.001
Chronic workload (CL)	0.000	0.001	0.969	1.000	0.999	1.001
AL - CL difference (coupled)	0.001	0.000	0.025	1.001	1.000	1.001
AL - CL difference (uncoupled)	0.000	0.000	0.107	1.000	1.000	1.001
Acute workload	0.001	0.000	0.011	1.001	1.000	1.001
CL < 1328	0.334	0.480	0.486	1.396	0.545	3.574
1329 < CL < 1522	0.216	0.387	0.576	1.241	0.582	2.648
1523 < CL < 1717	0.020	0.380	0.958	1.020	0.484	2.150
CL > 1718 (ref)	0.000	.	.	1.000	.	.
Acute workload	0.002	0.001	0.062	1.002	1.000	1.004
Chronic workload	0.001	0.001	0.570	1.001	0.998	1.003
AL * CL	0.000	0.000	0.223	1.000	1.000	1.000
Log(Acute workload)	1.159	0.376	0.002	3.188	1.53	6.65
Log(Chronic workload)	-1.012	0.777	0.193	0.364	0.079	1.67
Week to week AL difference	0.000	0.000	0.021	1.000	1.000	1.001
Chronic workload	0.000	0.001	0.936	1.000	0.999	1.001
Week to week AL difference	0.000	0.000	0.022	1.000	1.000	1.001
CL < 1328	-0.073	0.412	0.860	0.930	0.414	2.087
1329 < CL < 1522	-0.036	0.356	0.920	0.965	0.481	1.937
1523 < CL < 1717	-0.117	0.363	0.747	0.890	0.437	1.812
CL > 1718 (ref)	0.000	.	.	1.000	.	.
AL - CL difference (coupled)	0.001	0.000	0.017	1.001	1.000	1.001
Chronic workload	0.000	0.001	0.870	1.000	0.999	1.001
AL - CL difference (coupled)	0.001	0.000	0.019	1.001	1.000	1.001
CL < 1328	-0.093	0.401	0.817	0.911	0.416	1.998
1329 < CL < 1522	-0.042	0.353	0.905	0.959	0.480	1.914
1523 < CL < 1717	-0.127	0.363	0.726	0.881	0.432	1.794
CL > 1718 (ref)	0.000	.	.	1.000	.	.
AL - CL difference (uncoupled)	0.003	0.001	0.001	1.003	1.001	1.005
Chronic workload	0.000	0.000	0.285	1.000	1.000	1.001
AL-CL * CL	0.000	0.000	0.003	1.000	1.000	1.000
AL - CL difference (uncoupled)	0.000	0.000	0.101	1.000	1.000	1.001
CL < 1328	-0.057	0.403	0.888	0.945	0.429	2.081
1329 < CL < 1522	-0.010	0.356	0.977	0.990	0.493	1.987
1523 < CL < 1717	-0.110	0.364	0.764	0.896	0.439	1.831
CL > 1718 (ref)	0.000	.	.	1.000	.	.

Table 3. Parameters of various models estimated using the original acute:chronic workload ratio (ACWR from 2 to 4 weeks) and ACWRs created using fixed values for the chronic workload, for whole sample and players' data dichotomised in two groups based on the chronic load median value.

	B	Std. Error	Wald	Exp(B)	95% Confidence Interval	
			Sig.	OR	Lower	Upper
Original ACWR values						
ACWR 4 weeks	0.896	0.333	0.007	2.451	1.276	4.707
ACWR 3 weeks	1.121	0.296	0.000	3.069	1.718	5.485
ACWR 2 weeks	1.172	0.388	0.003	3.228	1.510	6.900
ACWR with fixed values						
Acute/500	0.221	0.100	0.027	1.247	1.025	1.517
Acute/1000	0.441	0.200	0.027	1.555	1.051	2.301
Acute/1510	0.667	0.302	0.027	1.948	1.078	3.520
Acute/2000	0.883	0.400	0.027	2.418	1.104	5.295
Acute/2500	1.104	0.500	0.027	3.015	1.132	8.032
Subgroup (chronic load > 1523 AU)						
Acute workload	0.000	0.000	0.529	1.000	1.000	1.001
ACWR 4 weeks	0.342	0.784	0.663	1.407	0.303	6.542
Acute/1510	0.316	0.502	0.529	1.372	0.513	3.670
Subgroup (chronic load < 1523 AU)						
Acute workload	0.001	0.000	0.001	1.001	1.000	1.001
ACWR 4 weeks	1.276	0.374	0.001	3.583	1.721	7.463
Acute/1510	1.066	0.292	0.000	2.903	1.639	5.143

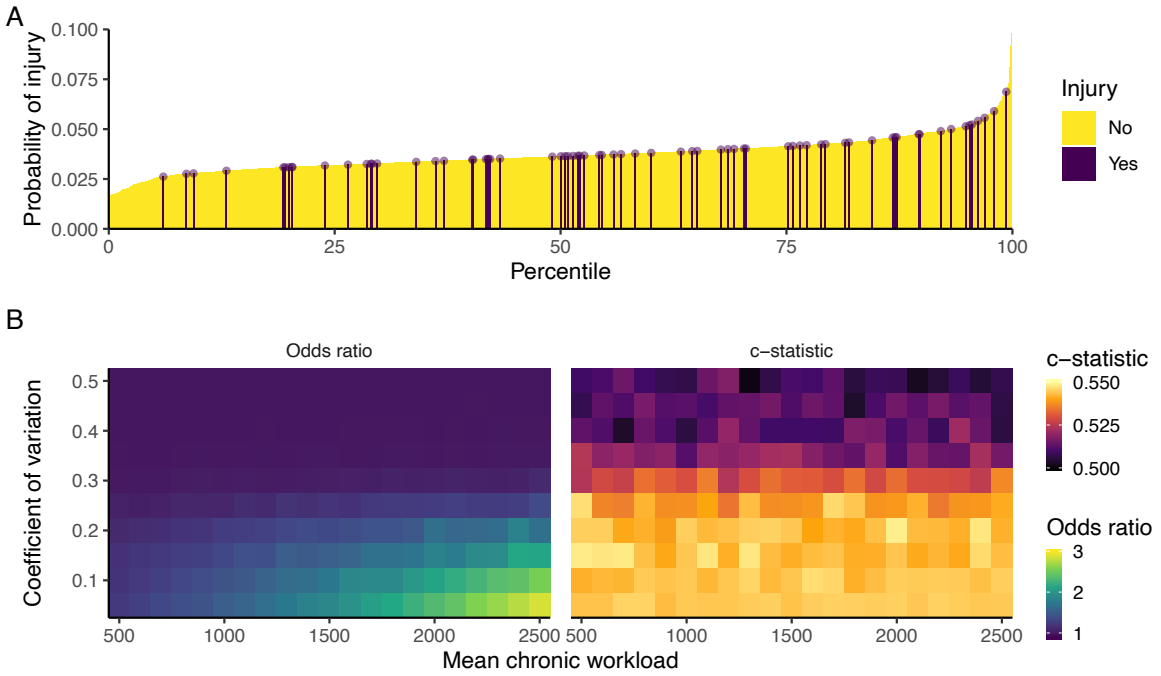

Figure 2. Results from models using true and random acute-to-chronic workload ratios. (**A**) fitted values (probability) for each *real* data point included in the model. Even though the acute-to-chronic workload ratio has a statistically significant odds ratio that is greater than 1, all probabilities of injury are low, and the probabilities of those who were injured do not clearly separate from those who were not. This indicates that the model is poorly calibrated and acute-to-chronic workload ratio does not contain predictive information, even when "tested" in the training dataset. (**B**) results for models that use random chronic workloads. Left, as mean chronic workload increases and coefficient of variation decreases, odds ratios increase—this is a basic statistical property of ratios. Right, *c*-statistics or AUCs from random chronic workload models are similar to that from the model that uses true chronic workload (cf. 0.574), and all are also similar to an intercept-only model (cf. 0.5).

Table 4. Differences between injured and uninjured players

Explanatory variables (AU)		N	Mean	SD	P level	95% Confidence Interval			Effect size (pooled SD)
						Mean difference	Lower	Higher	
Acute workload	Injured	70	1601	337	0.063	78	-4	161	0.18
	Uninjured	1849	1523	445					
Acute/1510	Injured	70	1.061	0.223	0.063	0.052	-0.003	0.107	0.18
	Uninjured	1849	1.009	0.295					
ACWR 4 weeks	Injured	70	1.085	0.219	0.010	0.071	0.018	0.125	0.26
	Uninjured	1849	1.014	0.271					
ACWR 3 weeks	Injured	70	1.092	0.210	0.002	0.081	0.030	0.132	0.31
	Uninjured	1849	1.011	0.259					
ACWR 2 weeks	Injured	70	1.083	0.229	0.009	0.076	0.020	0.131	0.31
	Uninjured	1849	1.008	0.244					
Subgroup (Chronic > 1523 AU)									
Acute Load	Injured	34	1758	290	0.612	26	-77	130	0.07
	Uninjured	926	1732	358					
ACWR 4 weeks	Injured	34	1.011	0.171	0.675	0.013	-0.048	0.074	0.07
	Uninjured	926	0.998	0.193					
Subgroup (Chronic < 1523 AU)									
Acute Load	Injured	36	1453	314	0.013	140	31	250	0.33
	Uninjured	923	1313	425					
ACWR 4 weeks	Injured	36	1.155	0.239	0.004	0.126	0.043	0.209	0.38
	Uninjured	923	1.029	0.331					

Table 5. Cross tabulation to show the reclassification of individual player data

High chronic group (> 1523 AU)		ACWR				
		< 0.87	**0.88-1.02**	**1.03-1.17**	**> 1.17**	**Total**
Acute load	**< 1261**	73	7	0	0	80
	1262-1542	37	164	0	0	201
	1543-1851	1	186	105	0	292
	>1852	0	51	264	72	387
	Total	111	408	369	72	960
	Injured	3	13	16	2	34

Low chronic group (< 1523 AU)		ACWR				
		< 0.87	**0.88-1.02**	**1.03-1.17**	**> 1.17**	**Total**
Acute load	**< 1261**	174	168	43	15	400
	1262-1542	0	112	131	36	279
	1543-1851	0	0	119	69	188
	>1852	0	0	2	90	92
	Total	174	280	295	210	959
	Injured	1	11	12	12*	36

*, n=3 from acute load <1261; n=1 from 1262-1542; n=8 from 1543-1851

Appendix 1.

Coefficients, p values, odd ratios (ExpB) with 95% confidence intervals for the odd ratios estimated using as explanatory variable the acute:chronic workload ratios calculated dividing each player acute workload by randomly generated chronic workload values having the same standard deviation than the original chronic workload (282 SD), lower (141 AU) or higher (423 AU).

Parameter	B	Sig	ExpB	Lower	Upper
From random CL with SD=282					
Random set 1	0.33	0.11	1.39	0.93	2.08
Random set 2	0.34	0.10	1.41	0.94	2.11
Random set 3	0.72	0.01	2.05	1.23	3.42
Random set 4	0.70	0.00	2.01	1.27	3.18
Random set 5	0.34	0.15	1.40	0.89	2.22
Random set 6	0.17	0.38	1.19	0.81	1.74
Random set 7	0.52	0.01	1.69	1.16	2.46
Random set 8	0.51	0.02	1.67	1.07	2.58
Random set 9	0.41	0.10	1.51	0.92	2.47
Random set 10	0.40	0.02	1.50	1.06	2.12
Random set 11	0.35	0.12	1.42	0.91	2.21
Random set 12	0.28	0.14	1.33	0.91	1.94
Random set 13	0.36	0.04	1.43	1.02	1.99
Random set 14	0.43	0.03	1.54	1.04	2.29
Random set 15	0.32	0.21	1.38	0.84	2.27
Random set 16	0.54	0.08	1.72	0.95	3.12
Random set 17	0.40	0.07	1.48	0.98	2.26
Random set 18	0.50	0.05	1.64	1.01	2.67
Random set 19	0.15	0.56	1.16	0.71	1.89
Random set 20	0.48	0.04	1.61	1.03	2.50
Random set 21	0.33	0.15	1.39	0.89	2.19
Random set 22	0.41	0.06	1.50	0.98	2.31
Random set 23	0.73	0.00	2.07	1.33	3.21
Random set 24	0.35	0.19	1.42	0.84	2.40
Random set 25	0.36	0.19	1.43	0.83	2.47
From random CL with SD=141					
Random set 1	0.66	0.01	1.94	1.22	3.07
Random set 2	0.77	0.01	2.15	1.24	3.75
Random set 3	0.66	0.02	1.93	1.10	3.37
Random set 4	0.51	0.07	1.66	0.96	2.88
Random set 5	0.52	0.09	1.68	0.92	3.04
Random set 6	0.66	0.03	1.93	1.05	3.56

Random set 7	0.53	0.07	1.69	0.97	2.96
Random set 8	0.67	0.01	1.96	1.18	3.26
Random set 9	0.65	0.02	1.91	1.11	3.28
Random set 10	0.59	0.05	1.81	1.01	3.25
Random set 11	0.59	0.05	1.81	1.01	3.25
Random set 12	0.61	0.02	1.84	1.12	3.03
Random set 13	0.50	0.08	1.64	0.94	2.85
Random set 14	0.83	0.00	2.30	1.35	3.91
Random set 15	0.50	0.07	1.65	0.96	2.83
Random set 16	0.35	0.18	1.42	0.85	2.40
Random set 17	0.71	0.01	2.03	1.17	3.54
Random set 18	0.78	0.01	2.17	1.24	3.81
Random set 19	0.46	0.09	1.58	0.93	2.70
Random set 20	0.99	0.00	2.70	1.48	4.92
From random CL with SD=423					
Random set 1	0.27	0.16	1.31	0.90	1.91
Random set 2	0.01	0.39	1.01	0.99	1.03
Random set 3	0.12	0.07	1.12	0.99	1.28
Random set 4	0.08	0.38	1.09	0.90	1.30
Random set 5	0.15	0.22	1.17	0.91	1.49
Random set 6	0.15	0.03	1.16	1.01	1.32
Random set 7	0.07	0.55	1.08	0.85	1.36
Random set 8	0.26	0.02	1.29	1.05	1.59
Random set 9	0.04	0.76	1.04	0.79	1.37
Random set 10	0.05	0.63	1.05	0.86	1.28
Random set 11	-0.12	0.31	0.89	0.71	1.12
Random set 12	0.04	0.76	1.04	0.82	1.32
Random set 13	0.12	0.36	1.13	0.87	1.47
Random set 14	0.24	0.11	1.28	0.95	1.72
Random set 15	0.14	0.31	1.15	0.88	1.50
Random set 16	0.23	0.19	1.26	0.89	1.77
Random set 17	0.24	0.06	1.27	0.99	1.63
Random set 18	0.26	0.03	1.30	1.03	1.63
Random set 19	0.01	0.90	1.01	0.82	1.25
Random set 20	0.17	0.45	1.18	0.76	1.84

Appendix 2.

Within player relations between acute workload and chronic workloads