PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Getting the most out of intensive longitudinal data: A methodological review of workload—injury studies
AUTHORS	Windt, Johann Ardern, CL; Gabbett, Tim; Khan, Karim; Cook, Chad; Sporer, Ben; Zumbo, Bruno

VERSION 1 – REVIEW

REVIEWER	Dr Catia Malvaso Research Associate, School of Psychology and School of Public Health, University of Adelaide, Australia
REVIEW RETURNED	13-Apr-2018

GENERAL COMMENTS	Overall this is a very detailed and insightful review. The authors are correct in stating that many reviews and quantitative syntheses do not take into account the methods and statistical approaches used in included studies. Examining the methodological features and choice of statistical analyses provides an important perspective into the quality of evidence in any given research field. Simply synthesising evidence does not always provide insight into the quality of that evidence – methodological reviews can fill this gap. I believe that there are a number of considerations that require attention before publication. Because the topic of workload-injury is not necessarily part of my expertise, some of my comments might not be directly relevant. I have focused my comments on my areas of expertise, namely the general flow and structure of the paper, methods and statistical analysis, with some questions that come to mind for a reviewer that is not an expert in the workload-injury field. Suggestions are delineated below according to the sections of the review. Further to these, I suggest the authors carefully proofread their manuscript as a number of minor wording issues, punctuation issues and errors were identified (e.g., page 4, para 3, use of '-'; page 6, para 2, 3 and 4, various instances of potential missing words; page 17, para 1, missing '(' etc. Abstract 1. Some of the information provided does not appear to fit within the subheadings. Some examples include: the first line of the objectives appears to a conclusion rather than an objective; generally the aims should appear in the objectives not in the methods sections as it stands currently. Method 2. Was there any consideration given to the competition level of included sports teams? I know this information is included in Table

	1, but some discussion might be required. For example, in football, an English premier league club's workload for their athletes would differ significantly from the expectations for a championship league team, to a league one or two team, despite all being professional teams. How would this influence associations between workload and injury? If the review was dominated by studies using data from premier league clubs (or equivalent) would this change any of the conclusions? The same would apply with different types of team sports. 3. In the article coding and description section, the authors state that 50 items of information were extracted from included papers. However, in the results section, the authors state that 500 criteria were coded. Did I miss something? Discussion 4. Overall, a lot of what is given in the discuss is probably better suited to results. For example, on page 24, the authors write about the ways assumptions were checked in the included papers. The majority of this is descriptive information, and there is very little interpretation or discussion (with the exception of the final paragraph in the section on page 25). I would recommend re-organising some of this information so that purely descriptive information is given in the results, and more space devoted to writing about the expectations and recommendations for work in this area to follow in the discussion section. 5. Some further discussion around the reason to control for known risk factors would be helpful. For example, while controlling for age might seem obvious, why would sex be important? What other risk factors (aside from previous injury) have been/should be included? 6. Some further discussions about complex systems approach and analyses is suggested. How would self-organising feature maps etc, help in this area? Would taking an approach similar to the in the field of epidemiology, for example directed acyclic graphs, be useful? What about machine learning (I know this is mentioned later in the paper, but some more detailed discussion about some of the most promising approaches would strengthen the paper). 7. Do the authors have any suggestions for the best approach to dealing with missing data in this area (rather than simply saying that researchers should account for missing data by whichever strategies)? Can you discuss methods for imputation that might be more or less relevant? For example, if studies commonly aggregate data or are looking at interactions in a regression analysis, then researchers should be carefully considering which imputation methods they use. Von Hippel, for example, recommends the transform-then-impute method. 8. Some further discussion about 'effect-measure modifiers' on page 27 would be helpful. Many people confuse interactions (or moderation) with effect modification. Could interaction and effect modification coincide in this area? What approach would be used in this instance? The example that follows was useful, but can be expanded upon further for clarity.

REVIEWER	William Riley National Institutes of Health, USA
REVIEW RETURNED	28-May-2018

GENERAL COMMENTS	This paper provides an excellent and well-grounded perspective on the statistical challenges of intensive longitudinal data (ILD) and how at least one area of research (sports injury) needs to make substantial improvements in statistical approaches that are better

	suited to these types of data. The methodological systematic review is well-described and appropriate. The summary of the results is clear and well-organized. The list of considerations (highlighted in the first paragraph and described in more detail in the discussion) for ILD data is an excellent rubric for investigators using these types of data. There are a few minor weaknesses to address: 1. It is difficult to judge the statistical appropriateness of ILD research conducted over many years. Those in the early period of using these data could be more easily "forgiven" for aggregating data over time and using statistical approaches that, in some cases, were not developed and available when these were published. Minimally, acknowledging this in the discussion would be helpful, but it also may be useful to analyze their data over time (ILD for an ILD review) and determine if there have been improvements in how these data are analyzed as the field has evolved. 2. The review and discussion seems a bit GEE heavy. This may be the result of GEE used more frequently in this literature, but mixed models have greater flexibility, and Hedeker's hierarchical mixed models have been developed specifically for ILD data. The authors do mention time varying effect models (TVEMs) and other more recent approaches, but it may be useful for the reader to have less of a list of what's been done previously (much of it inadequate to the challenges of ILD) and more of what can and should be done - including a list of the pros and cons of our current ILD approaches. While it is understandable that the authors may not want to make definitive recommendations, the field could use guidance on the statistical approaches and how they handle the various challenges. 3. In abstract, they say "Lisa Collins" but it's Linda Collins.

REVIEWER	Ana Maria Staicu North Carolina State University, USA
REVIEW RETURNED	19-Jun-2018

GENERAL COMMENTS	The paper discusses whether/how athletes training and competition workload relate to injury risk. The manuscript is written as a review of the published work on this topic with the intention to evaluate whether the methods comply to the three-fold alignment - theoretical model, temporal design and stats model - describe in of Collins (2006), and provide recommendation for future investigation. I find the topic interesting and timely. However, I consider that the authors fell short of their goals and find their contribution rather modest. The authors' objective is not trivial due to the various ways the injury was defined as well as the ways the workload was measured in the past publications. They do a good job and clarifying these definitions; however I find the presentation of the methods involved and really reviewing past work on this topic superficial. The reader would appreciate a critical review of the past published papers on this topic that offers a critical eye or deep insight into the methodology used.

	I appreciated the short survey of the papers published on this topic, which is presented in Table 1. In the current form it is hard to follow the information in the table; I recommend that the authors order the publications first by sport and then by year of publication. The data source table is disappointing; stating that 15 papers use data from "previous journal article in field of study" add close to no information. If this is the intention, then the authors should summarize the table in 1-2 lines. A more informative table would be to cite the data sources. When summarizing the published work I recommend to start with the dominant method and then discuss other approaches that have a less frequent usage.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 – Dr Catia Malvaso

REVIEWER COMMENT: Overall this is a very detailed and insightful review. The authors are correct in stating that many reviews and quantitative syntheses do not take into account the methods and statistical approaches used in included studies. Examining the methodological features and choice of statistical analyses provides an important perspective into the quality of evidence in any given research field. Simply synthesising evidence does not always provide insight into the quality of that evidence – methodological reviews can fill this gap.

AUTHOR RESPONSE: We thank Dr. Malvaso for her time in thoroughly reviewing our manuscript – her insight and recommendations have served to improve our revised manuscript. We also want to thank you for your kind words regarding our review – it was our intention to help fill the gap you discuss by providing some insight into the quality of evidence underlying this field, and we are confident that our revision better accomplishes that goal.

REVIEWER COMMENT:
I believe that there are a number of considerations that require attention before publication. Because the topic of workload-injury is not necessarily part of my expertise, some of my comments might not be directly relevant. I have focused my comments on my areas of expertise, namely the general flow and structure of the paper, methods and statistical analysis, with some questions that come to mind for a reviewer who is not an expert in the workload-injury field. Suggestions are delineated below according to the sections of the review. Further to these, I suggest the authors carefully proofread their manuscript as a number of minor wording issues, punctuation issues and errors were identified (e.g., page 4, para 3, use of ‘-’; page 6, para 2, 3 and 4, various instances of potential missing words; page 17, para 1, missing ‘(’ etc.

AUTHOR RESPONSE AND MANUSCRIPT CHANGES:

Thank you for highlighting these grammatical/punctuation details. We have corrected all the changes you have highlighted. Further, we have carefully proofread the revised manuscript to correct other minor wording and punctuation issues throughout.

Abstract

REVIEWER COMMENT: 1. Some of the information provided does not appear to fit within the subheadings. Some examples include: the first line of the objectives appears to a conclusion rather than an objective; generally the aims should appear in the objectives not in the methods sections as it stands currently.

AUTHOR RESPONSE AND MANUSCRIPT CHANGE: Thank you. Our revised abstract now begins as follows.
Objectives: To systematically identify and qualitatively review the statistical approaches used in prospective cohort studies of team sports that reported ILD (>20 observations per athlete) and examined the relationship between athletic workloads and injuries. Since longitudinal research can be improved by aligning the (1) theoretical model, (2) temporal design, and (3) statistical approach, we reviewed the statistical approaches used in these studies to evaluate how closely they aligned these three components.
Design: Methodological review.
Methods: After finding 6 systematic reviews and 1 consensus statement in our systematic search, we extracted 34 original prospective cohort studies of team sports that reported ILD (>20 observations per athlete) and examined the relationship between athletic workloads and injuries. Using Prof. Linda Collins' three-part framework of aligning the theoretical model, temporal design, and statistical approach, we qualitatively assessed how well the statistical approaches aligned with the intensive longitudinal nature of the data, and with the underlying theoretical model. Finally, we discussed the implications of each statistical approach and provide recommendations for future research.

Methods

REVIEWER COMMENT:
2. Was there any consideration given to the competition level of included sports teams? I know this information is included in Table 1, but some discussion might be required. For example, in football, an English premier league club's workload for their athletes would differ significantly from the expectations for a championship league team, to a league one or two team, despite all being professional teams. How would this influence associations between workload and injury? If the review was dominated by studies using data from premier league clubs (or equivalent) would this change any of the conclusions? The same would apply with different types of team sports.

AUTHOR RESPONSE:
The specific sporting context for each of these studies is very important if we are to accurately interpret, summarize, and synthesize the workload—injury studies. As you've highlighted, all 'Professional' soccer teams are not the same. In addition to your point of workload, other factors such as, travel demands and health professional support vary greatly and may effect a synthesis of the evidence.
The benefit of our current methodological review is that the principles of choosing a statistical approach that aligns with the longitudinal nature of the data should be quite consistent across different levels of play and different sports. The major limitation that we believe could come into play based on competition level is that lower competition teams may not have the budget to collect as much data, and therefore may be limited in multivariable analyses, something we highlighted in our discussion.

MANUSCRIPT CHANGE:
We have added the following paragraph to our review limitations to address these potential impacts. "In some instances, authors' analytical choices may have been attributable to factors outside of statistical considerations. For example, in lower level competitions, or in organizations with lower budgets, it may not have been feasible to collect multiple variables longitudinally with the available equipment or staff. In these types of instances, authors would be unable to employ a multifactorial approach, instead of choosing not to use one. Such external factors may have influenced the findings of this methodological review."

REVIEWER COMMENT:
3. In the article coding and description section, the authors state that 50 items of information were extracted from included papers. However, in the results section, the authors state that 500 criteria were coded. Did I miss something?

AUTHOR RESPONSE:
We agree that additional clarification would be helpful. There were 50 items from each of the included papers, so a total of 500 items were coded by each coder in the in the first 10 papers. This provided the data for our analysis of agreement– 98%. We have reworded this sentence to be clearer about this process and calculation.

MANUSCRIPT CHANGE:
"Thirty-four articles were included in this methodological review (Appendix 1). In the first 10 articles coded by both reviewers, there were 10 discrepancies out of 500 total coded entries (10 papers x 50 items/paper) which gave us 98% agreement between reviewers."

Discussion
REVIEWER COMMENT:
4. Overall, a lot of what is given in the discussion is probably better suited to results. For example, on page 24, the authors write about the ways assumptions were checked in the included papers. The majority of this is descriptive information, and there is very little interpretation or discussion (with the exception of the final paragraph in the section on page 25). I would recommend re-organising some of this information so that purely descriptive information is given in the results, and more space devoted to writing about the expectations and recommendations for work in this area to follow in the discussion section.

AUTHOR RESPONSE AND MANUSCRIPT CHANGES:
We revised the manuscript to rearrange information accordingly. We moved the descriptive information regarding assumption checking to the Results as you suggested. Thank you for helping us better focus the Discussion to critical analysis and recommendations for future research in this revision.

REVIEWER COMMENT:
5. Some further discussion around the reason to control for known risk factors would be helpful. For example, while controlling for age might seem obvious, why would sex be important? What other risk factors (aside from previous injury) have been/should be included?

AUTHOR RESPONSE:
Thank you for pointing out that further discussion of why controlling for these risk factors is important, using age and sex as an example. You alerted us that this part of the Discussion would allow us to explain interaction, effect-measure modification, and mediation (motivated by your comment #8). We have added the paragraph below into our 'Consideration #1 – Theoretical theme – multifactorial aetiology' section.

MANUSCRIPT CHANGE:
"Including known risk factors in workload—injury investigations is important from an aetiological perspective in at least two ways. First, failing to control for known risk factors may mean that key confounding variables are not included in the analysis and the relationship between workloads and injury are spurious. For example, women have a 2-6 times higher risk of ACL injury in soccer than their male counterparts [60,61]. If a study included both male and female soccer players and did not account for sex in the analysis, then differences in workload may be spuriously correlated with injury rates if male and female players performed varying levels of workload. Depending on the injury type and sporting group, previous injury, age, sex, physiological and/or biomechanical variables may all be important to include.

Secondly, by including additional risk factors into the analysis, the investigator may be able to identify moderation or effect-measure modification to better understand how risk factors and workload jointly contribute to injury risk [62,63]. As a reminder, there are subtle, but important differences between mediation, moderation, and effect measure modification that will influence analytical choices [64,65]. Effect modification occurs when the effect of a treatment or condition (e.g. a given workload demand), differs among different athlete groups. Interaction (or moderation), although similar, examines the joint effect of two or more variables on an outcome. Finally, mediation is concerned with the pathway of exposure to a given outcome, and what are potentially intermediate variables. Previously identified risk factors may aetiologically relate to workload in each of these three ways and may be explored through different modelling strategies."

REVIEWER COMMENT
6. Some further discussions about complex systems approach and analyses is suggested. How would self-organising feature maps etc, help in this area? Would taking an approach similar to the in the field of epidemiology, for example directed acyclic graphs, be useful? What about machine learning (I know this is mentioned later in the paper, but some more detailed discussion about some of the most promising approaches would strengthen the paper).

AUTHOR RESPONSE:
In response to your comment here and Dr. Riley's comment about an expanded 'prescriptive' section to the paper, we have concluded the Discussion with a section about 'Future directions for ILD analysis'. We have done our best to balance a review of previous approaches and prescriptive recommendations for the future within a reasonable word limit.

We devoted a paragraph to discussing the benefits of time-to-event models and mixed modelling approaches (both under-utilized) and acknowledged other approaches that could also be beneficial. Some of these future possibilities extend into the complex system approaches, where we mention self-organizing feature maps, DAGs, machine learning and others. This guides readers and opens the door for a more detailed manuscript detailing these approaches in the future

MANUSCRIPT CHANGE:
Changes to the manuscript can be seen in the final section of the Discussion.

Researchers in the sports medicine field should be encouraged that the increased availability of ILD may improve understanding of athletes' fluctuating injury risks – as articulated by their theoretical models. More advanced statistical techniques for longitudinal data are increasingly being developed and implemented across disciplines. This will enable sports medicine researchers to more accurately answer their theory-driven questions by taking advantage of the benefits of ILD. To capitalise on this understanding, researchers must choose statistical models that most closely align with their theory and that address longitudinal data challenges. Generalised estimating equations, a Cox proportional hazards model, a multilevel logistic model, and a frailty model were the 4 analyses that most closely approached this alignment within our included papers. However, there remains some clear room for improvement in the future.

First, although mixed modelling was only used in one study, these forms of analyses have inherent values over GEE methods and have been recommended for this reason [101]. Because of sample structure, mixed models prevent false positive associations and have an applied correction method that increases the power of the analysis [102]; a finding that is useful with the commonly smaller samples. Mixed models also carry a less stringent missing data assumption (missing at random) when compared with GEEs (missing completely at random). Further, whereas GEEs require the correlation structure to be chosen by the researcher (which may be wrong), mixed models model the correlation structure so that it can be investigated. Finally, GEEs assume a constant effect across all individuals in the model, while mixed models allow for individual level effects and for differentiating these individual effects.

To borrow an example from another field and demonstrate the flexibility and utility of mixed effect models, Russell et al. used daily stressor values from students during their first 3 college years to demonstrate that students consumed more alcohol on high-stress days than low-stress days (within-person fixed effect) [103]. However, a significant random effect between students suggested that some students experienced this increase in alcohol consumption, while others did not. Finally, those students with a tendency to increase alcohol consumption with stressors were more likely to have drinking-related problems in their 4th year [103]. For more information on multilevel/mixed effect models for longitudinal analysis, readers are referred to a other helpful resources [1,28,75,104,105].

Time-to-event models are another family of statistical models that have become a very common in clinical research articles – reported in 61% of original articles in the New England Journal of Medicine in 2004-2005 [106] – but were used infrequently within our included articles. Notably, these models answer a different research question – when does an event occur? These approaches can account for many of the ILD challenges [107–109]. Time-to-event models account for censoring, can incorporate time-varying exposures, time-varying effect measure modifiers, and time-varying changes in injury status, and may be used to control for competing risks [107]. As with other modelling techniques, the appropriate number of events per variable has been investigated, and at least 5-10 events per variable are recommended for these types of models to prevent sparse data bias [110]. As long as this and other model assumptions are met, more advanced time-to-event models may be a valuable tool for researchers analysing ILD [77,111,112].

Lastly, computational modelling methods, which involves computer simulation has both pros and cons where modelling injuries. On one hand, they may provide insight on the best ways to model certain predictor variables[113], and open the door to more complex systems modelling (e.g. agent-based modelling) [91]. Though they show promise, such simulation studies are based on artificially generated data and must be interpreted carefully [114].

More analytical approaches are available for ILD, but a full discussion of each of these is beyond the scope of this paper. For the interested reader, functional data analysis [115], machine learning approaches [92,95], time series analysis [116], and time-varying effect models [117] all show promise.

Such analyses and others for ILD can be found in Walls and Schafer's landmark ILD textbook [1], and more recently, in the work of Bolger and Laurenceau [104].

REVIEWER COMMENT:
7. Do the authors have any suggestions for the best approach to dealing with missing data in this area (rather than simply saying that researchers should account for missing data by whichever strategies)? Can you discuss methods for imputation that might be more or less relevant? For example, if studies commonly aggregate data or are looking at interactions in a regression analysis, then researchers should be carefully considering which imputation methods they use. Von Hippel, for example, recommends the transform-then-impute method.

AUTHOR RESPONSE:
Thank you for allowing us this opportunity. We expanded our discussion of missing data in the sports medicine/sport science field and longitudinal data analysis in general. To include a more prescriptive component to the discussion of missing data, we conclude with a paragraph that suggests approaches for missing data with ILD, drawing from von Hippel and others.

MANUSCRIPT CHANGE:
"Dealing with missing and unbalanced data is a near certainty when collecting ILD, and is common in applied workload-monitoring settings [71]. Such missing data decreases statistical power and increases bias, and may be missing at completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). When analysing aggregated data or using analyses that require balanced data, strategies may include complete-case analysis, last observation carried forward, or various imputation methods [72,73]. Multiple imputation methods, of which there are many, involves replacing missing values with values imputed from the observed data and is preferred over single imputation. Finally, if interactions are included in regression analyses, the transform-then-impute method has been recommended [74].

However, these missing data approaches are not recommended for longitudinal analyses, since researchers have statistical analyses that are robust to missing and unbalanced data at their disposal [75]. Statistically, four types of analyses used in this review are robust to missing and unbalanced data – Cox proportional hazards models, GEEs, multilevel models, and frailty models, where all observations can be included in the analysis, and athletes can have different numbers of observations. Since mixed/multilevel models have less stringent assumptions for missing data (i.e. missing at random) than generalized estimating equations (i.e. missing completely at random), they have been suggested over GEEs [75].

While the statistical concerns related to unbalanced data may be addressed with these analyses, missing data may also affect derived variables, which are common in workload—injury research. These derived variables include rolling workload averages (e.g. one-week, 'acute', workloads, four-week average, 'chronic', workloads, etc.) [33,41], 'monotony' (average weekly workload divided by the standard deviation of that workload) or 'strain' (the monotony multiplied by the average weekly workload) [30]. Since these measures are all calculated from workloads accumulated over time, failing to estimate workloads for these missing sessions (that end up being treated as '0' workload days) means inferences from these derived measures may be underestimated and unreliable. Few authors discussed how they handled missing data. In these instances – it is important that researchers report how they accounted for missing data, whether they be strategies employed in the past – e.g. full team average values (Bowen et al., 2016), weekly individual averages (Brink et al., 2010), player specific per-minute values by time played (Colby et al., 2014) – or whether through other advanced imputation methods recommended for ILD [72,74]."

REVIEWER COMMENT:
8. Some further discussion about ‘effect-measure modifiers’ on page 27 would be helpful. Many people confuse interactions (or moderation) with effect modification. Could interaction and effect modification coincide in this area? What approach would be used in this instance? The example that follows was useful, but can be expanded upon further for clarity.

AUTHOR RESPONSE:
Explicitly outlining the differences in moderation and effect modification is a helpful addition to the manuscript and we have clarified the Discussion in this regard in two ways referencing the two sources below. First, under Consideration #1 – multifactorial aetiology, we discuss moderation & effect measure modification as ways to understand risk factors in aetiology. We are also more explicit in our example near the end of the Discussion (formerly page 27).

Corraini, Priscila, Morten Olsen, Lars Pedersen, Olaf M Dekkers, and Jan P Vandenbroucke. “Effect Modification, Interaction and Mediation: An Overview of Theoretical Insights for Clinical Investigators.” Clinical Epidemiology 9 (June 8, 2017): 331–38. https://doi.org/10.2147/CLEP.S129728.

VanderWeele, Tyler J. “On the Distinction Between Interaction and Effect Modification:” Epidemiology 20, no. 6 (November 2009): 863–71. https://doi.org/10.1097/EDE.0b013e3181ba333c.

MANUSCRIPT CHANGE:
In the ‘Consideration #1 section, our revised paragraph reads:
“Secondly, by including additional risk factors into the analysis, the investigator may be able to identify moderation or effect-measure modification to better understand how risk factors and workload jointly contribute to injury risk [62,63]. As a reminder, there are subtle, but important differences between mediation, moderation, and effect measure modification that will influence analytical choices [64,65]. Effect modification occurs when the effect of a treatment or condition (e.g. a given workload demand), differs among different athlete groups. Interaction (or moderation), although similar, examines the joint effect of two or more variables on an outcome. Finally, mediation is concerned with the pathway of exposure to a given outcome, and what are potentially intermediate variables. Previously identified risk factors may aetiologically relate to workload in each of these three ways and may be explored through different modelling strategies.”
In discussing the developments in longitudinal analysis on what was page 27, the paragraph now reads:

"Mediation, effect measure modification, and interaction/moderation are all causal models which may also contribute to aetiological frameworks [98]. We recently proposed that traditional intrinsic and extrinsic risk factors may act as moderators or effect measure modifiers of the workload—injury association [62]. If that is true, the most appropriate statistical model would include workload measures as the independent variable of interest, and incorporate other risk factors such that these causal models can be investigated, whether by stratifying effects across different levels of these risk factors, or including an interaction term within regression [63]. While no included articles performed such an analysis, recent studies (not included in this review because it was published after our search) have started to adopt these approaches [94,99,100]. For example, Møller et al. used a frailty model with weekly workload fluctuations (decrease or <20% increase, 20-60% increase, and >60% increase) as the primary predictor variable in a frailty model. Known shoulder risk factors were treated as ‘effect measure modifiers’, so the model was stratified based on the presence or absence of a given risk factor (e.g. scapular dyskinesis) [65]. In so doing, the researchers used a statistical tool (Component #3) that addressed all the challenges inherent to longitudinal data (Component #2), conducting a multifactorial analysis that clearly differentiated both within- and between-athlete injury risk – key aspects of the theoretical model (Component #1).

Reviewer 2 – William Riley

REVIEWER 2, COMMENT 1: This paper provides an excellent and well-grounded perspective on the statistical challenges of intensive longitudinal data (ILD) and how at least one area of research (sports injury) needs to make substantial improvements in statistical approaches that are better suited to these types of data. The methodological systematic review is well-described and appropriate. The summary of the results is clear and well-organized. The list of considerations (highlighted in the first paragraph and described in more detail in the discussion) for ILD data is an excellent rubric for investigators using these types of data. There are a few minor weaknesses to address:

AUTHOR RESPONSE: We thank Dr. Riley for his time in reviewing our manuscript, and his kind words regarding the review. We are pleased that he finds the structure helpful and overall message of the paper clear, relevant, and grounded. With the revisions made in response to Dr. Riley's comments, along with those of the other reviewers, we are confident that we strengthened the manuscript.

REVIEWER 2, COMMENT 2:
1. It is difficult to judge the statistical appropriateness of ILD research conducted over many years. Those in the early period of using these data could be more easily "forgiven" for aggregating data over time and using statistical approaches that, in some cases, were not developed and available when these were published. Minimally, acknowledging this in the discussion would be helpful, but it also may be useful to analyze their data over time (ILD for an ILD review) and determine if there have been improvements in how these data are analyzed as the field has evolved.

AUTHOR RESPONSE:
This is an important consideration that addresses a key characteristic of scientific inquiry – that research questions, methods, and knowledge all improve and evolve over time. Since the original submission of the review, several original prospective cohort studies have been published, and few have used the aggregated approaches which we discuss as 'limited' in the first submission of this review. We have taken your advice and highlighted it in a new Discussion section.

MANUSCRIPT CHANGES:
Instead of discussing evolving methods over time as a sub-section of our 'limitations' section, we have created a new sub-heading entitled "Longitudinal improvements in ILD analysis". Within this section we discuss the evolution in longitudinal analysis methods and aetiological models over time. Finally, we have included an analysis of these data over time by providing an 'average score' for papers over time based on our scoring criteria. The first two paragraphs of this section now read:

"Methods and statistical analyses evolve over time, as with all scientific inquiry. Therefore, it is possible that we were a little unfair to some earlier papers. For example, researchers may have chosen analyses that aligned with 'their' theoretical model at the time, not what is considered the most current theoretical model. However, most papers were published since 2010 – the dynamic, recursive aetiology model was introduced in 2007, and the multifactorial nature of injury risk has been highlighted since 1994 (Meeuwisse, 1994). As complex systems approaches are the most recently proposed theoretical model (Bittencourt et al., 2016, Hulme et al., 2015), it is not surprising that none of the included articles analysed the data within this type of framework, with the first analysis of its kind in sport injury research only appearing recently [91]. Further, some techniques for longitudinal data analysis have been developed and grown in popularity recently, so researchers may not have been aware of alternative approaches at the time of their studies.

As more statistical methods are developed and refined for longitudinal data analysis, researchers will continue to gain awareness and skills with these analyses and their implementation is likely to become more common. Some evidence for that progression can be seen in this review. If we were to assign a ‘method’ score to each analytical approach outlined in Table 1, assigning 0 for each red box, 0.5 for each yellow box, and 1 for each green box (e.g. correlation would score 0, while generalized estimating equations would score a 3.5), and then assign that score to each paper in the study, we could obtain a rough estimate of whether analytical approaches were improving over time. Breaking the papers roughly into four periods, the ‘average score’ for papers up to 2005 (n = 6) is 1.6, papers between 2006 and 2010 (n = 7) score an average of 1.9, papers between 2011-2015 (n=11) score 1.7, and papers since 2016 (n = 10) score an average of 2.3. Moreover, since the search for this current review was conducted, there have been promising developments in the sports medicine field and a continued improvement in longitudinal analysis. Recent publications have applied statistical models that more appropriately take advantage of the strengths inherent to ILD, and better align with the theoretical frameworks [92–97].”

REVIEWER 2, COMMENT 3:
2. The review and discussion seems a bit GEE heavy. This may be the result of GEE used more frequently in this literature, but mixed models have greater flexibility, and Hedeker's hierarchical mixed models have been developed specifically for ILD data. The authors do mention time varying effect models (TVEMs) and other more recent approaches, but it may be useful for the reader to have less of a list of what's been done previously (much of it inadequate to the challenges of ILD) and more of what can and should be done - including a list of the pros and cons of our current ILD approaches. While it is understandable that the authors may not want to make definitive recommendations, the field could use guidance on the statistical approaches and how they handle the various challenges.

AUTHOR RESPONSE:
Indeed, the GEE-heavy nature of the discussion resulted from that method’s prevalence in this literature. Most papers published in the last couple months in the workload—injury sphere have also used GEEs, as they have become the most common modelling approach in the field at this time. In our first submission, we focused on what had been done in the field. We take the point and were more prescriptive – not solely descriptive. In this revision, we have devoted a section to ‘Future Directions for ILD Analysis.

In this revision we aimed to incorporate additional discussion of future recommendations in a concise manner. We extended the benefits of mixed modelling approaches and time-to-event models (which have largely been under-utilized in the field), and alert the reader to other promising methods. Rearranging these sections means that we conclude the Discussion by describing how longitudinal analysis has developed over the course of our review period, and where it is likely to head.

MANUSCRIPT CHANGE:
We have now reformatted the paper’s discussion to address both of Dr. Riley’s primary recommendations. Along with the new sub-heading after the ‘review limitations’ that discusses ‘longitudinal improvements in ILD analysis’, we have revised the final section of our paper to address ‘Future directions for ILD analysis’ instead of ‘Bright spots and future directions’. That section reads as follows:
“Researchers in the sports medicine field should be encouraged that the increased availability of ILD may improve understanding of athletes’ fluctuating injury risks – as articulated by their theoretical models. To capitalise on this understanding, researchers must choose statistical models that most

closely align with their theory, and that address longitudinal data challenges. Generalised estimating equations, a Cox proportional hazards model, a multilevel logistic model, and a frailty model were the 4 analyses that most closely approached this alignment within our included papers. However, these are not the only options available to researchers.

More advanced statistical techniques for longitudinal data are increasingly being developed and implemented across disciplines. This will enable sports medicine researchers to answer their theory-driven questions while taking advantage of the benefits of ILD.
Worth noting is that several more applicable methods were not used in any of the studies. Mixed modelling was only used in one study, and these forms of analyses have inherent values over GEE methods. Because of sample structure, mixed models prevent false positive associations and have an applied correction method that increases the power of the analysis [92]; a finding that is useful with the commonly smaller samples. Causal mediation analyses allow an understanding of the roles of intermediate variables that lie in the causal path between the predictor and outcome, and allows the researcher to focus on both the longitudinal mediating and primary predictor variables that are associated with the targeted outcome [93]. Other forms of time-to-event (TTE) analyses, such as advanced survival methods, evaluate both the outcome of interest (whether or not an event occurred), but also when that event occurred [94–96]. Software to perform these methods are often lacking. Lastly, computational modelling methods, which involves computer simulation has both pros and cons where modelling injuries. Whereas these simulation models can provide multiple potential scenarios, they have also been known to misrepresent data and present false information.

Although beyond the scope of this paper, there are other approaches which have potential to positively impact longitudinal data analysis. Functional data analysis [97], machine learning approaches [82,85],time series analysis [98], and time-varying effect models [99] all show promise. An overview of some of these methods can be found in Walls' and Schafer's landmark textbook for ILD [1], and more recently, in the work of Bolger and Laurenceau [100].

Finally, we believe there is an exciting opportunity for applied researchers and statisticians to collaborate moving forward. As the field continues to progress to more advanced analytical approaches that may better suit ILD, the need for collaboration with statisticians will be vital. In our included papers, few researchers referenced methodological or statistical references to justify their analytical approaches. In some instances, this may be attributable to using common, relatively simple analyses – one likely doesn't expect a citation for a t-test. Where such references existed, they were often to previous papers in the field, not statistical sources. In future longitudinal analyses, we encourage researchers to partner with a statistician, psychometrician, biostatistician, etc., working together to facilitate more optimal, theory-driven approaches [101]."

REVIEWER 2, COMMENT 4.
3. In abstract, they say "Lisa Collins" but it's Linda Collins.

AUTHOR RESPONSE AND MANUSCRIPT CHANGE:
We are so sorry! Apologies and we have corrected Professor Linda Collins' name in the abstract.

Reviewer 3 - Ana Maria Staicu

REVIEWER COMMENT:
The paper discusses whether/how athletes training and competition workload relate to injury risk. The manuscript is written as a review of the published work on this topic with the intention to evaluate

whether the methods comply to the three-fold alignment - theoretical model, temporal design and stats model - describe in of Collins (2006), and provide recommendation for future investigation. I find the topic interesting and timely. However, I consider that the authors fell short of their goals and find their contribution rather modest.
The authors' objective is not trivial due to the various ways the injury was defined as well as the ways the workload was measured in the past publications. They do a good job and clarifying these definitions; however I find the presentation of the methods involved and really reviewing past work on this topic superficial. The reader would appreciate a critical review of the past published papers on this topic that offers a critical eye or deep insight into the methodology used.

AUTHOR RESPONSE:
We would like to thank Dr. Staicu for her review and constructive comments on our manuscript. We believe that our responses to all the reviewers' comments has notably improved our manuscript and that our revised submission provides a deeper insight for our audience and a proposed direction forward, while maintaining an accessible and readable structure through the topic.

REVIEWER COMMENT:
I appreciated the short survey of the papers published on this topic, which is presented in Table 1. In the current form it is hard to follow the information in the table; I recommend that the authors order the publications first by sport and then by year of publication.

AUTHOR RESPONSE AND MANUSCRIPT CHANGE:
We agree that this arrangement would improve the original Table that was organized by Author Name and thank the reviewer for this suggestion. We have re-formatted Table 1 so that it is now sorted by Sport first, and then by year of publication.

REVIEWER COMMENT:
The data source table is disappointing; stating that 15 papers use data from "previous journal article in field of study" add close to no information. If this is the intention, then the authors should summarize the table in 1-2 lines. A more informative table would be to cite the data sources.

AUTHOR RESPONSE:
We understand Prof. Staicu's comment regarding this table and the limited value that it adds. Fundamentally, the primary takeaway from this section is that few of the included papers referenced statistical sources when explaining or justifying their analyses. Expanding the table to identify all the original cohort studies that the authors referenced would add information, but likely not contribute to our intention in that particular section. Therefore, as you suggested, we have removed Table 5, and summarized the section in the text instead. Moreover, we expanded on this intention of the section in the final section of our discussion. In response to the other reviewers' recommendations for more discussion about future directions/recommended analysis, we have included a new Table in the discussion (Table 6) outlining some future analytical approaches for ILD, their pros/cons, and key references.
Finally, we now conclude our discussion with the exciting opportunity for applied researchers and statisticians to collaborate more closely as these approaches become more common.

MANUSCRIPT CHANGES:
We removed table 5, and the section now reads in text only:
"Fifteen of the included articles (44%) did not cite any sources to support their analytical choices. Of those that did, most (n=14) cited previous literature in the sports medicine field. Eight articles

referenced statistics or methodology articles, 4 cited Will Hopkins' website (www.sportssci.org), and 3 cited a statistical textbook."
Near the end of the discussion we have a paragraph that now reads:
We believe intensive longitudinal data provide an exciting opportunity for applied researchers and statisticians to collaborate moving forward. As the field continues to progress to more advanced analytical approaches that may better suit ILD, the need for collaboration with statisticians will be vital. In our included papers, few researchers referenced methodological or statistical references to justify their analytical approaches. In some instances, this may be attributable to using common, relatively simple analyses – one likely does not expect a citation for a t-test. Where such references existed, they were often to previous papers in the field, not statistical sources. In future longitudinal analyses, we encourage researchers to partner with a statistician, psychometrician, epidemiologist, biostatistician, etc [118]. Such fruitful collaborations may lead to statistical approaches that take full advantage of intensive longitudinal data by aligning theory, data collection and statistical analyses as seamlessly as possible.

REVIEWER COMMENT:
When summarizing the published work I recommend to start with the dominant method and then discuss other approaches that have a less frequent usage.

AUTHOR RESPONSE:
We agree with Prof. Staicu that the structure and flow of the discussion is important for the ease and understanding of the reader and have modified these paragraphs accordingly.

MANUSCRIPT CHANGES:
We have modified the flow of the results which we believe have improved the revised version. In accordance with Prof. Staicu's recommendation, we have reordered the results section 'typical uses of statistical tools' to begin with the most common method (regression approaches) and proceed through to the least frequent methods. We have also restructured table 4 so that it proceeds through this same order, from the most frequent to least frequent use.

VERSION 2 – REVIEW

REVIEWER	Dr Catia G Malvaso University of Adelaide, Australia
REVIEW RETURNED	01-Aug-2018

GENERAL COMMENTS	The authors have done a thorough job in addressing all of my comments. The manuscript has been significantly improved.