BMJ Open

Getting the most out of intensive longitudinal data: A methodological review of workload—injury studies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022626
Article Type:	Research
Date Submitted by the Author:	13-Mar-2018
Complete List of Authors:	Windt, Johann; University of British Columbia, Experimental Medicine Ardern, CL; Linkoping University, Division of Physiotherapy Gabbett, Tim; Gabbett Performance Solutions; University of Southern Queensland Khan, Karim; The University of British Columbia, Centre for Hip Health & Mobility Cook, Chad; Duke University, Department of Orthopedics Sporer, Ben; University of British Columbia, Family Practice Zumbo, Bruno; University of British Columbia, Measurement, Evaluation, and Research Methodology Program
Keywords:	STATISTICS & RESEARCH METHODS, PREVENTIVE MEDICINE, SPORTS MEDICINE

Running Title: ALIGNING THEORY, DATA COLLECTION, AND STATISTICAL ANALYSES IN SPORTS MEDICINE RESEARCH USING LONGITUDINAL DATA

Getting the most out of intensive longitudinal data: A methodological review of workload—injury studies

Johann Windt,[1,2,3] Clare L Ardern,[4,5] Tim J. Gabbett,[6,7] Karim M. Khan,[1] Chad E. Cook,[8] Ben Sporer,[9,10] Bruno D. Zumbo[11]

[1] *Experimental Medicine Program, University of British Columbia, Vancouver, Canada*

[2] *United States Olympic Committee, Colorado Springs, United States of America*

[3] *United States Coalition for the Prevention of Illness and Injury in Sport, Colorado Springs, United States of America*

[4] *Division of Physiotherapy, Linköping University, Linköping, Sweden*

[5] *School of Allied Health, La Trobe University, Melbourne, Australia*

[6] *Gabbett Performance Solutions, Brisbane, Australia*

[7] *University of Southern Queensland, Institute for Resilient Regions, Ipswich, Australia*

[8] *Department of Orthopaedics, Duke University, Durham, United States of America*

[9] *Department of Family Practice, University of British Columbia*

[10] *Vancouver Whitecaps Football Club, Vancouver, British Columbia*

[11] *Measurement, Evaluation, and Research Methodology Program, University of British Columbia, Vancouver, Canada*

Corresponding Author:

Johann Windt, MSc, CSCS (ORCID ID: 0000-0001-5517-2609)

2635 Laurel Street, Vancouver, BC V5Z 1M9.

Email: johannwindt@gmail.com

Key Words: Methodology, workloads, training load, athletic injury, statistics

Word Count: *7085*

Objectives: The quality of longitudinal research can be improved by authors aligning the (1) theoretical model, (2) temporal design, and (3) statistical approach. Intensive longitudinal data (ILD) are increasingly common and pose methodological/statistical challenges that increase the need to align these three components. The field of sports medicine provides an example where many prospective cohort studies include ILD; a specific example that provides substrate to evaluate the statistical approaches is in studies of the association between athletic workloads and injury. Therefore, we reviewed the statistical approaches used in prospective cohort studies that used ILD to examine the association between athletic workloads and injury.

Design: Methodological review.

Methods: We systematically identified, and qualitatively reviewed, the statistical approaches used in 34 prospective cohort studies of team sports that reported ILD (>20 observations per athlete) and examined the relationship between athletic workloads and injuries. Using Lisa Collins' three-part framework, our primary aim was to assess how well the statistical approaches aligned with the intensive longitudinal nature of the data, and with the underlying theoretical model.

Results: Statistical methods such as correlations, t-tests, and simple linear/logistic regression were commonly used. However, these methods did not adequately address the (1) themes of theoretical models underlying workloads and injury, nor the (2) temporal design challenges (ILD). Time-to-event analyses (e.g. Cox Proportional Hazards and frailty models), generalized estimating equations and multilevel modelling are better-suited for ILD, but these were utilized in fewer than a 30% of the studies (n=10).

Conclusions: Rapidly accelerating availability of intensive longitudinal data is the norm in many fields of health care delivery and thus health research. These data present an opportunity to better address research questions, especially when appropriate statistical analyses are chosen.

STRENGTHS AND LIMITATIONS

- As intensive longitudinal data are becoming increasingly common across disciplines, catalysed by technological advances, this methodological review provides researchers with a number of considerations when determining how they will analyse these data.
- Whereas systematic reviews provide a quantitative synthesis of research findings, they do not account for the statistical approaches used in the original studies. Therefore, methodological reviews like this one fill an important void in the literature to highlight key shortcomings and ways forward from a methodological and statistical perspective.
- By choosing a homogenous group of papers – prospective cohort studies in team sports that collected intensive longitudinal data – we were able to focus more directly on the statistical analyses that authors employed.
- It was beyond the scope of this review to list every challenge posed by intensive longitudinal data, and we are not exhaustive in our discussion of different analyses and their capacity to handle the challenges that we did highlight.

INTRODUCTION

Intensive longitudinal data (ILD) are being collected more frequently in various research areas [1], catalysed by technological advancements that simplify data collection and analysis [2]. By collecting data repeatedly on the same participants, researchers are enabled to answer more detailed research questions, particularly regarding phenomena that change or fluctuate over time. However, arriving at these answers requires researchers to overcome the challenges of analysing ILD, which include: (1) the dependencies created by repeated measures, (2) missing/unbalanced data, (3) separating between- and within-person effects, (4) time-varying and time-invariant (stable) factors, and (5) specifying the role of time/temporality [3].

The field of exercise and sports medicine provides one specific example which can illustrate principles that apply to the use of intensive longitudinal data broadly. In the field of sports performance, technological advances mean that a plethora of physiological, psychological and physical data are conveniently available from athletes [4,5]. As one example, of professional 48 football clubs that responded to a survey on player monitoring, 100% reported collecting daily global positioning system (GPS) and heart rate (HR) data [6].

One research question that has gained a great deal of interest in the last decade is how athletes' training and competition workloads relate to injury risk. Since athletes' training and injury risk continually varies over time – many researchers have used prospective cohort studies to collect and analyse ILD to answer this question [7]. There is moderate evidence from systematic reviews and an International Olympic Committee (IOC) consensus statement suggesting a positive relationship between injury rates and high training workloads, increased risk of injury with low workloads, and a pronounced increase in injury risk associated with rapid workload increases [7–11]. However, such systematic reviews do not consider the statistical approaches used in included studies [12]. Choosing the wrong statistical analysis or poorly implementing an otherwise correct one (e.g. violating statistical assumptions) can bias results and create false conclusions. Even a perfectly performed systematic review cannot compensate for poorly designed, or poorly analysed studies [13].

Longitudinal data analysis is most effective when the chosen statistical approach aligns with the frequency of data collection and with the theoretical model underpinning the research question (See Box 1) [14]. Therefore, we used this lens to evaluate whether the statistical models employed in prospective cohort studies using ILD to investigate the relation between athletic workloads and injury were optimal. There were 3 aims: (1) to summarise researchers' data collection, methodological, statistical, and reporting practices [12,15]; (2) to evaluate the degree to which the adopted statistical analyses fit within Collins' three-fold alignment; and (3) to provide recommendations for future investigations in the field.

Box 1: Theoretical model – temporal design – analytical model.

In a landmark, highly-cited paper, Professor Linda Collins described how aligning the (1) theoretical model (subject matter theory), (2) temporal design (data collection strategy/timing), and (3) statistical model (analytical strategy) is crucial when analysing longitudinal data [14]. *For example, if researchers (1) theorize that a given physiological variable fluctuates every hour, (2) data must be collected at least on an hourly basis. If researchers measure participants once a day, they will miss virtually all the hourly fluctuations that their theories predict. Once researchers have collected their hourly data, they should (3) select a statistical strategy that enables them to examine the relationship between these fluctuations and the outcome of interest. As Collins noted, perfect alignment of these 3 components may not be possible, but it provides researchers a target, and readers a lens through which longitudinal research can be evaluated.*

METHODS

Article selection

We systematically searched the literature (MEDLINE, CINAHL, SPORTDISCUS, PsychInfo, and EMBASE) (December 10, 2016) to identify systematic reviews and consensus statements that investigated the relationship between workloads and athletic injuries, with the aim of extracting all original articles included in these reviews that met our inclusion criteria. A summary of the systematic search and article selection process is described in Appendix 1 (Table A1 and Figure A1), and the full systematic search is available from the authors.

A priori, we operationally defined 'workload' as either external – the amount of work completed by the athlete (e.g. distance run, hours completed, etc.) or internal – the athlete's response to a given external workload (e.g. session rating of perceived exertion, heart-rate based measures, etc.). We acknowledge that athlete self-reported measures often evaluate how athletes are handling the demands of training and may be referred to as 'internal' load measures, but we considered these perceptual wellbeing measures as a distinct step from quantifying athletes' internal or external workloads [16]. Athletic injuries have been diversely defined in the literature, so we operationally defined athletic injury as any article that reported measuring 'injury', regardless of their specific definition (e.g. time loss, medical attention, etc).

Two authors (JW + TG) screened the titles/abstracts of the systematic reviews. Where necessary, the full texts of the reviews retrieved in the search. A total of 6 systematic reviews [7–10,17,18] and 1 consensus statement [11] were identified that included at least 1 article meeting the inclusion criteria.

We extracted and reviewed the full texts of all the original studies included (n=279) in these 7 papers. For our analysis included all the original articles that met the following criteria:

1) Original articles were prospective cohort studies that examined the relationship between at least 1 measure of internal or external workload (as defined above) and athletic injury. Since theoretical models describe the recursive nature of injury risk with each training or competition exposure, workloads had to be continually monitored and include both training and match workloads for the same athletes. Although some athletes may have entered or left the group during the study period (e.g. through retirement or trades to other teams) the same team/group of athletes had to be followed throughout the study period, as opposed to repeated cross-sectional snapshots of different
2) Articles collected intensive longitudinal data. We defined intensive longitudinal data as >20 observations per athlete [14].
3) Articles studied team sport athletes. We chose team sports because (1) there are high amounts of ILD collected in applied team sport settings [6], and (2) the majority of workload—injury studies are in team

sport athletes (Jones, 2016). Military populations and individual sports (e.g. distance running) were excluded due to the differences in task requirements and operating environment.

Patient and public involvement

As a methodological review, there was no patient or public involvement in this current investigation.

Article coding and description

To describe the methodological, statistical, and reporting approaches utilized in each article, two authors (JW + CA) reviewed all the included papers and extracted 50 items of information, including publication year, journal, variable operationalization (e.g. internal vs. external load measures, injury definition, etc), methodological approaches, statistical analyses implemented, reported findings, and more. To ensure consistency between coders, 10 articles were randomly selected and coded independently by both reviewers. Discrepancies were discussed by the two coders and an additional 5 articles were randomly selected and coded. The remaining articles were coded by JW and checked by CA.

Assessing statistical models' alignment with Collins' threefold framework

To evaluate the statistical approaches used in this field, we first identified the key themes and challenges within the theoretical models and temporal design features within the workload—injury field, then developed a qualitative assessment to evaluate the statistical approaches.

Collins Component 1: The theoretical models that underpin athletic workloads and injury risk (in brief)

Briefly, we identified at least 3 key elements of athletic injury aetiology models. First, *sports injuries are multifactorial* [19–21]. Aetiology models since 1994 have all explained between-athlete differences in injury risk by identifying a host of 'internal' (e.g. athlete characteristics, psychological wellbeing, previous injury) and 'external' (e.g. opponent behaviour, playing surface) risk factors. More recently, Meeuwisse et al.'s dynamic recursive model [22] and the workload—injury aetiology model [23], have highlighted the recurrent nature of injury risk, meaning each athlete's injury risk (i.e. within-athlete risk) also fluctuates continually as they train or compete in their sport (Figure 1). Thus, a second theme is that *injury risk differs between- and within-athletes*. Finally, more recent injury aetiology models have highlighted *injury risk as a complex, dynamic system* (Figure 2)

[24,25]. Complex systems, as in weather forecasting or biological systems [26,27], possess many key features, including an open-system, inherent non-linearity between variables and outcomes, recursive loops where the system output becomes the new system input, self-organization where regular patterns (risk profiles) may emerge for given outcomes (emergent pattern), and uncertainty [24].

--- INSERT FIGURE 1 HERE ---

Figure 1 – The workload—injury aetiology model. Key features include the multifactorial nature of injury, between- and within- athlete differences in risk, and a recursive loop.

--- INSERT FIGURE 2 HERE ---

Figure 2 – Complex systems model of athletic injury. Web of determinants are shown for an ACL injury in basketball players (A), and in a ballet dancer (B)

Collins' 2nd Component – Temporal design / data collection

The theoretical models relating workloads and injury illustrate a continuously fluctuating injury risk, with many variables that influence risk on a daily or weekly basis [22–24]. Thus, if researchers want to investigate the association between workloads and injuries, these data must be collected frequently enough to observe changes in these variables as they occur (temporal design). With technological advances, athletes' physiological, psychological and physical variables are now often collected on a daily, weekly or monthly basis, along with ongoing injury surveillance data [4,5]. Therefore, in the workload—injury field, the theoretical models (injury aetiology models that describe regular fluctuation in workloads and injury risk) and the temporal design (frequent, often daily, data collection) are often well-aligned, especially in prospective cohort studies using ILD. This leaves us to consider only Professor Collins' third component – the statistical model – aligns with these first two.

Collins' Component 3 – Statistical model

From the theoretical aetiology models underpinning the workload—injury association, we highlighted three key themes to consider when choosing a statistical model: (1) injury risk is multifactorial, (2) between-athlete and within-athlete differences in injury risk fluctuate regularly, and (3) injury risk may be considered a complex, dynamic system.

From a temporal design perspective, intensive longitudinal data (ILD) are necessary to address these key themes, but they also carry at least 5 challenges that influence the choice of the statistical model.

1) Differentiating between- and within-person effects.
2) ILD include time-varying variables (e.g. workloads), and may also incorporate stable (time-invariant) variables (e.g., sex).
3) The 'dependency' created by repeated measurements of the same individuals violates the assumption of 'independence' common to many traditional analyses [28,29].
4) Almost all longitudinal datasets have missing or unbalanced data [14]. Researchers must learn how best to deal with missing data (e.g. imputation methods), and where appropriate, choose analyses that are sufficiently robust to handle missing/unbalanced data.
5) Longitudinal data analysis require researchers to consider the role of time in their analysis [3].

Evaluating statistical approaches

We deliberately tried to align components 1 and 2 of Collins' framework by describing the theoretical models underpinning the workload—injury association and only including articles that had a temporal design characterised by ILD. To review whether statistical approaches aligned with these two components, two authors (JW + BZ) qualitatively assessed whether the statistical models, as employed in the included studies, (1) were multifactorial, (2) differentiated between- and within-athlete differences in injury risk, and (3) analyzed the data as a dynamic system – the three themes highlighted in the theoretical framework. From the temporal design, the same two authors evaluated whether the statistical analyses (4) included both time-varying and time-invariant variables, (5) were robust to missing/unbalanced data, 6) addressed the dependencies created by repeated measures, and (7) incorporated time into the analysis.

Data synthesis approach

We first describe the characteristics of the included articles, then present our qualitative assessment of how well the various statistical approaches fit within Collins' framework.

RESULTS

Thirty-four articles were included in this methodological review (Appendix 1). In the first 10 articles coded by both reviewers, 500 criteria were coded, with 10 discrepancies between reviewers (98% agreement). No item had more than 2 discrepancies. Of the 250 study criteria in the second set of 5 articles coded by both reviewers, there were 8 discrepancies (97% agreement).

Included articles were published from 2003 – 2016, with 78% of the studies published since 2010. Sports studied included rugby league (n = 10), soccer (n = 7), Australian football (n = 6), cricket (n = 5), rugby union (n = 2), multiple sports (n = 1), and basketball, handball, and volleyball (n = 1 each). Studies included an average of 96 athletes (median = 46), ranging from 12 [30] to 502 athletes [31]. The observation period for these cohort studies ranged from 14 weeks [32] to 6 years [33]. Most studies investigated male athletes (n = 30), with 2 studies on female athletes, and 2 on both sexes. Table 1 summarizes the basic characteristics of the included articles, while the full data extraction table is available from the authors upon request.

Data collection

Injury definitions

Injury definitions varied across articles, with exact wording outlined in the Online Supplementary Appendix. In Table 2, we have categorised the definitions into more discrete injury categories (and subcategories) in accordance with recognized consensus statements [34].Where studies used multiple injury definitions, we categorized them according to the definition used for the primary analysis.

Reference	Journal	Study Length	Sport	Level	Sex	Age
(Anderson et al., 2003) [30]	*JSCR*	21 weeks	Basketball	Sub-elite competitive	Female	18-22
(Arnason et al., 2004) [35]	*AJSM*	1 season	Soccer	Professional	Male	24 (16-38)
(Bowen et al., 2016) [36]	*BJSM*	2 seasons	Soccer	Elite Youth Players	Male	17.3 ± 0.9
(Bresciani et al., 2010) [37]	*Eur J Sport Sci*	1 season (40 weeks)	Handball	Elite	Male	20.1 (2.5)
(Brink et al., 2010) [38]	*JSCR*	2 years	Soccer	Elite Youth Players	Male	16.5 ± 1.2
(Brooks et al., 2008) [31]	*J Sports Sci*	2 seasons	Rugby Union	Professional	Male	Not reported
(Clausen et al., 2014) [39]	*AJSM*	1 season	Soccer	Recreational / Adolsecent	Female	15-18
(Colby et al., 2014) [40]	*JSCR*	1 season	AFL	Professional	Male	25.1 ± 3.4
(Cross et al., 2015) [41]	*IJSPP*	1 season	Rugby Union	Professional	Male	Not reported
(Dennis et al., 2004) [42]	*J Sci Med Sport*	1 season	Cricket	Professional	Male	25.2 (range = 21-34)
(Dennis et al., 2003) [43]	*J Sci Med Sport*	2 seasons	Cricket	Professional	Male	27 (range 18 - 38)
(Dennis et al., 2005) [44]	*BJSM*	1 season	Cricket	Sub-elite	Male	14.7 years
(Duhig et al., 2016) [45]	*BJSM*	2 seasons	AFL	Professional	Male	22.2 (3.4)
(Ehrmann et al., 2015) [46]	*JSCR*	1 season	Soccer	Professional	Male	25.7 (5.1) Range 18-33
(Gabbett, 2010) [47]	*JSCR*	4 years	Rugby league	Professional	Male	23.7 (3.8)
(Tim J. Gabbett, 2004) [48]	*J Sports Sci*	1 season (2001)	Rugby league	Semi-professional	Male	Not reported
(Tim J. Gabbett, 2004) [49]	*BJSM*	3 years (2001-2003)	Rugby league	Sub-elite	Male	22.9 (20.7-25.1) in 2001 19.6 (18.4 to 20.8) in 2002 21.5 (19.6 to 23.4) in 2003
(Gabbett and Jenkins, 2011) [50]	*J Sci Med Sport*	4 years	Rugby league	Professional	Male	23.3 (3.8)
(Gabbett and Domrow, 2007) [51]	*J Sports Sci*	2 seasons	Rugby league	"Sub-elite"	Male	21.4 ± 5.1
(Gabbett and Ullah, 2012) [52]	*JSCR*	1 season	Rugby league	Professional	Male	23.6 (3.8)
(Hulin et al., 2014) [33]	*BJSM*	6 years	Cricket	Professional	Male	26 ± 5
(Hulin et al., 2016a) [53]	*BJSM*	2 seasons	Rugby league	Professional	Male	24.8 (3.4)
(Hulin et al., 2016b) [53]	*BJSM*	2 seasons	Rugby league	Professional	Male	23.4 (3.5)
(Killen et al., 2010) [32]	*JSCR*	14 weeks	Rugby league	Professional	Male	17-32
(Malisoux et al., 2013) [54]	*J Sci Med Sport*	41 weeks	Varied (individual sports, racquet sports, team sports)	Competitive (high-school)	Both	14.1 (12-19)
(Mallo and Dellal, 2012) [55]	*J Sports Med Phys Fitness*	2 seasons	Soccer	Professional	Male	21.4
(Murray et al., 2016b) [56]	*Scand J Med Sci Sports*	2 seasons	AFL	Professional	Male	23 (4)
(Murray et al., 2016a) [57]	*IJSPP*	1 season	AFL	Professional	Male	23.1 (3.7)
(Owen et al., 2015) [58]	*J Sci Med Sport*	1 season	AFL	Elite	Male	22 (2.9)
(Rogalski et al., 2013) [59]	*BJSM*	1 season	Cricket	Elite	Male	24.4 (3.9). Range 18-32
(Saw et al., 2011)[60]	*J Sci Med Sport*	15 weeks	AFL	Elite	Male	23.4 (3.8)
(Veugelers et al., 2015)[61]	*Scand J Med Sci Sports*	4 years	Volleyball	Elite high school volleyball	Both (69 M, 72 F)	16.8 ± 0.8
(Visnes and Bahr, 2013) [62]	*BJSM*	1 season	Rugby league	Elite	Male	25 ± 3
(Windt et al., 2016) [68]	*JSCR*	2 seasons	Soccer	Professional	Male	26.8 years

Table 1: Summary of included workload—injury investigations

Table 2: Broad injury definitions used in workload—injury investigations

Injury Definition	N
Time-Loss	
All time-loss	13
Match time-loss	2
Non-contact time-loss	7
Non-contact match time-loss	1
Medical Attention	
Medical attention	7
Player-reported pain, soreness, or discomfort	1
Non-contact medical attention injuries	1
Clinical diagnosis of jumper's knee	1
Other	
Injury scale on the Recovery-Stress Questionnaire for Athletes (REST-Q)	1

Subsequent or recurrent injuries

Of the 34 articles, 30 did not define or include subsequent or recurrent injuries. Of those that explicitly addressed subsequent injuries, two defined these injuries as those occurring at the same time and occurring by the same mechanism [55,63]. Two articles explicitly stated that they only considered time until first injury, meaning no injuries were subsequent or recurrent [43,44].

Workload definitions

Workload variables varied widely across articles and are summarized in Table 3. For a more detailed description of each article's load measures, see the Online Supplementary Appendix. Many articles used workload metrics to derive additional variables from workload distribution over time (e.g. monotony, strain, acute:chronic workload ratios).

Table 3: Independent variables used in workload—injury investigations

Workload Measure	**N**
Internal	
- sRPE	15
- Heart rate zones	2
External	
- Balls bowled or pitched	5
- GPS / accelerometry	10
- Hours	6

* If articles included more than one type of workload variable they are counted more than once. sRPE scores could be the original Foster scale or modified. GPS – Global positioning system. sRPE – Session-rating of perceived exertion (calculated as the product of session intensity on a 1-10 Borg Scale and activity duration in minutes).

Measurement frequency

Most included articles (n=32) collected workload data at every session that athletes completed, while 2 studies recorded workload on a weekly basis [39,61].

Handling missing data

Twenty-three of the 34 articles (67%) did not report any strategies for missing data. Of those that did, 5 used listwise or casewise deletion, and 6 used estimation. Estimation methods for players missing data included techniques such as: using the full team average values for the drills a player completed [36], using an individual's mean weekly value [38], and multiplying player's pre-season per-minute match data by the number of minutes they played in a match [40].

Statistical analysis and reporting in included articles

Data binning/aggregation

Although 32 articles collected daily workload measurements, many aggregated data for analysis. Most (n = 16) summed workload metrics for a total or average weekly workload. Three studies aggregated workload data for the entire year, 3 aggregated data into season periods, 2 aggregated data monthly, and 3 used multiple aggregation strategies.

Analysis methods

Table 4 summarises the statistical practices of applied researchers investigating the relationship between workload and injury. Although some studies had analysed other primary or secondary objectives, we recorded only the analyses used to investigate the workload – injury relationship.

Table 4: The number of studies using various statistical analysis techniques

Analytical Method	**N**
Correlation	
- Pearson	9
- Spearman	1
T-tests	
- Paired and independent samples	4
- Independent samples only	2
Chi-square tests	1
Repeated measures ANOVA (one or two-way)	5
Relative risk/rate ratio*	8
Regression modelling	
- Logistic	
○ Regular	10
○ Generalised Estimating Equation	5
○ Multilevel	1
- Linear	
○ Regular	2
- Poisson	
○ Generalised Estimating Equation**	1
- Multinomial regression	
○ Regular	1
- Cox proportional hazards model	1
- Frailty model	1

If articles used more than one statistical method to analyse workload and injury, they are included more than once in the table. We only report analyses used to analyse workload—injury associations, not other analyses reported in the articles (e.g. ANOVA to test for differences in total workloads at separate times of the season).

* Relative risk here refers to the use of RR as a primary analysis based on risks in different categorical groups, not as an effect estimated from another model. For example, comparing risks among different load groups like Hulin et al., (2014, 2016, 2016) are counted here, whereas Gabbett and Ullah (2012) derived RR from their frailty model, and Clausen et al. (2014) derived RR from their Poisson model, but neither are included in the count for RR.

** Clausen et al. (2014) also report fitting multilevel models, but do not report any of the results – presenting only their GEE findings in results and discussion sections.

Typical uses of statistical tools

Most studies that used correlations (7/10) measured the association between weekly or monthly workloads and injury incidence at the team level. Of those that used correlation at the individual level, two compared the amount of pre-season sessions with the amount of in-season sessions completed [23,56], while the final compared workload with injury operationalised as a numerical score on the injury subscale of the Recovery-Stress Questionnaire for Athletes (REST-Q) [37].

Group comparisons were sometimes made using t-tests, ANOVAs or chi-square analyses. Typically, unpaired t-tests compared workload variables (e.g. mean sessions/week) between athletes who sustained an injury during the year, to those who did not [43,44]. Paired t-tests and repeated measures ANOVAs (one- or two-way) were most often used to compare the workloads of the same athletes at different time periods. For example, workloads in an 'injury block' (like the week preceding an injury), were compared with non-injury blocks, like other weeks in the season [43,59], or the 4 weeks preceding the injury block [54].

Relative risk approaches were used in two primary ways. First, workload categories were established for the entire year, like cricket bowlers who averaged <2, 2-2.99, 3-3.99, 4-4.99 or >5 days between bowling sessions up until an injury, or for the entire year if they didn't sustain an injury (Dennis, 2003). Risks were calculated as the number of injuries/number of athletes in a given group, and relative risks were calculated to compare across groups [59]. In the second approach, athletes contributed exposures on a weekly basis, and thus contributed to multiple workload classifications. In this case, the likelihood/risk was the number of injuries/number of weekly player exposures to that workload category (Bowen et al., 2016; Hulin et al., 2014, 2016).

As an overall category, regression approaches were the dominant statistical tool (22/34 studies), with multiple types of regression used. The most common approach was logistic regression (binary injury status as the outcome variable), independently or jointly modelling workload variables as independent variables. Generalised estimating equations were used to account for the clustering of observations within players and were used very similarly to simple logistic regression approaches.

Justifications for statistical approaches

Fifteen of the included articles (44%) did not cite any sources to support their analytical choices (Table 5). Of those that did, most (n=14) cited previous literature in the sports medicine field (Table 5).

Table 5: Sources cited to justify analyses used in workload—injury investigations

Citation Source	N
No references to justify approach	15
Previous journal article in field of study	14
Statistics or methodology article	8
Will Hopkins' website (www.sportsci.org/)	4
Statistical textbook	3

* Note – if articles cited another source for reason other than analysis (e.g. defining injury incidence), those citations are not included here.

Addressing analysis assumptions

More than half (n = 20) the included articles did not report on the assumptions underlying their statistical analyses. Among those that did report on analysis assumptions, checks included checks for normality, collinearity of predictor variables in regression analyses [41], sphericity for repeated measures ANOVA [37], overdispersion [39], or correlation structures for generalised estimating equations [60].

Alignment of authors' statistical models with theoretical model and temporal design challenges

In Table 6 (a more detailed table – Table A2 - is available in Appendix 2), we qualitatively evaluated whether the statistical approaches chosen by the authors in our current review effectively addressed the key themes/challenges presented by the theoretical model and the temporal design (intensive longitudinal data). This table is meant as an analytical tool to guide the reader through the discussion and highlight the themes/challenges of the theoretical model and temporal design, as well as the strengths/weaknesses of the statistical tools used in our included studies. The table is arranged such that the challenges/themes form columns, while the statistical tools form rows. Readers can thus follow a row to see how well a given statistical tool addressed key challenges as used by researchers in our included articles, or they can choose a challenge and follow the column down to see which analyses were used in a way that addressed that challenge adequately. The rows are ordered according to their qualitative 'score'. As one proceeds down the rows, the statistical tools can address more of the temporal design and theoretical model challenges.

We caution the reader that (1) not every possible statistical tool is included in the table, only those used in at least 1 article in our review, and (2) the evaluation is based on whether researchers of our included papers used a test in a way that addresses a given challenge, not necessarily whether the test is capable of being used in a way that meets that challenge. For example, a logistic regression analysis conducted using a generalised estimating equation framework can include multiple explanatory/predictor variables – thereby allowing for a multifactorial model). However, some authors used GEEs in this way but only used one predictor variable [42,51], so in this case the GEE did not address the multifactorial theme.

DISCUSSION

By design, the theoretical models underpinning the workload—injury field and the temporal design (ILD) were aligned in all the included articles, but common statistical approaches varied in how adequately they addressed the key themes needed to align them with the other two components.

Consideration #1 – Theoretical theme – multifactorial aetiology

Aetiological models of the last 2 decades have all highlighted the multifactorial nature of athletic injury, with both internal and external risk factors at play [19,21]. With the growing body of research highlighting the association of workloads to injury, one challenge is how to address the effects of workloads while incorporating known risk factors. Few articles in this review incorporated previously identified risk factors and workload into the same analysis. In some instances, the choice of analyses prevented this from being an option. For example, simple analyses like t-tests, correlations, and chi-square tests do not allow for multiple variables to be included. In other instances, the statistical approaches allowed a multifactorial approach (e.g. generalised estimating equations) but researchers opted to focus on the effects of workloads in isolation [42,51].

Table 6: Evaluation of the degree to which authors' use of statistical tools addressed theoretical and temporal design challenges

		Themes of theoretical model			Themes of temporal design - intensive longitudinal data			
Method	n	Multifactorial aetiology	Between and Within-Athlete Differences	Complex System	Includes Time-Varying and Time-Invariant Variables	Missing/ Unbalanced Data*	Repeated Measure Dependency	Incorporates Time into the Analysis
Correlation (Pearson and Spearman)	10	X	X	X	X	X	X	X
Unpaired t-test	6	X	X	X	X	X	X	X
Chi-square tests	1	X	X	X	X	X	X	X
Relative risk calculations	8	O	X	X	X	X	X	X
Regression (logistic, linear, multinomial)	13	O	X	X	X	X	X	X
Paired t-test	2	X	X	X	X	X	✓	✓
Repeated measures ANOVA (one or two-way)	5	O	O	X	O	X	✓	✓
Cox proportional hazards model	1	✓	X	X	X	✓	✓	✓
Generalised Estimating Equations (Poisson and logistic)	6	O	X	X	O	✓	✓	O
Multilevel modeling	2	✓	✓	X	✓	✓	✓	X
Frailty model	1	✓	✓	X	✓	✓	✓	✓

Qualitative assessment performed on a three-tiered scale. An 'X' (red formatting) means that none of the authors using this tool adequately addressed that specific challenge. In some cases, this may be because the statistical model was unable to address it, and other times it may be because of the way they used it. An 'O' (yellow formatting) indicates that some authors addressed that challenge while others did not. This generally happened when the statistical tool could address that challenge but the authors sometimes chose not to use it in that way. A '✓' (green formatting) indicates that all authors using this statistical tool addressed that challenge adequately. *Missing/unbalanced data here is that caused by intensive longitudinal data – meaning a different number of observations for each athlete during the observation period, some of which may be missing.

Other statistical approaches that allow multivariable analyses enabled researchers to examine the effects of workloads while controlling for known risk factors. Malisoux et al. (2013) used a Cox-proportional hazards model to control for age and sex while examining the effects of average training volume and intensity. Gabbett and Ullah's frailty model (2012) incorporated previous injury – a commonly cited injury risk factor – into the evaluation of the influence of different GPS workloads on injury risk. When investigating multifactorial phenomena, these types of statistical approaches that enable multiple explanatory variables provide a more appropriate option.

Consideration #2 – Theoretical theme - between and within-athlete differences

One of the primary benefits of ILD is that it enables researchers (when using certain analyses) to differentiate within-person and between-person effects [3]. In the sports medicine field, this would correspond to researchers asking, (1) why are some athletes robust against injury (between-person inquiry) while others are 'injury-prone'? and on the other hand, (2) at what point is a given athlete (within-person inquiry) more likely to sustain an injury? The simpler statistical approaches used by researchers in our included studies (correlation, t-tests, ANOVAs, regular regression) are limited in the number of variables they can include, and consequently cannot differentiate risk between- and within-athletes. Tests of group differences (independent sample t-tests and one-way ANOVAs) only differentiate between athletes (e.g. injured vs. uninjured), while repeated measures tests (repeated measures ANOVA and paired t-tests) only examine within-athlete differences (e.g. loads preceding injuries vs. loads during non-injury weeks).

Generalised estimating equations (GEE) were a common statistical choice used to deal with some of the longitudinal data challenges. However, although they account for the clustering within-persons, they assume the effects of predictor variables are constant across all athletes [64]. Simple Cox proportional hazards models (Malisoux et al. 2013) are common in survival analyses, but do not differentiate between- and within-person effects [65].

Only two statistical tools were used in a way that examined between- and within-athlete differences in injury risk. The frailty model by Gabbett and Ullah (2012) models each athlete as a random effect with a given frailty. The multilevel model by Windt et al., (2016) incorporated athlete-level variables (age, position, pre-season sessions) and observation-level variables (weekly workload measures). In the latter case, athletes' weekly distances did not affect their risk of injury in the subsequent week (OR = 0.82 for 1 standard deviation increase, 95% CI = 0.55 to 1.21) – a within-athlete inquiry. However, controlling for weekly distance and the proportion of distance at high speeds, athletes who had completed more pre-season training had significantly reduced odds of injury (OR = 0.83 for each 10 pre-season sessions, 95% CI = 0.70 to 0.99) – a between-athlete inquiry. These two examples highlight that certain analyses carry a distinct advantage of allowing researchers to tease out differences both between- and within- study participants – an important consideration when such differences are of interest.

Consideration #3 – Theoretical theme - injury risk as a complex dynamic system

Complex systems are defined, among other things, by the interaction between multiple internal and external variables that interact to produce an outcome. Simple analyses (t-tests, correlations), which cannot incorporate multiple variables, cannot examine the interaction between multiple factors. However, even other traditional analyses which are more effective in handling the challenges of longitudinal data (e.g. generalised estimating equations, Cox-proportional hazards models) were not used to incorporate non-linear interactions between predictor variables.

As the most recent 'theme' underling athletic injury aetiology, heuristic reviews on injury aetiology have increasingly presented as complex systems models [24,66]. However, none of the analyses included in our review analysed intensive longitudinal workload—injury data with statistical analyses that fit within a complex systems framework. This lack of research may reflect the fact that the suggestion that injury aetiology fits within a dynamic, complex systems framework is still relatively 'new'. It remains to be seen whether a complex systems approach and the analyses recommended in such reviews (e.g. self-organising feature maps, classification and regression trees, agent based models, etc.) are more effective for evaluating the association between workloads and injury [24].

Consideration #4 – ILD challenge – including time-varying and time-invariant variables

Tying back to the theoretical model of workloads and injury, some relevant factors may be relatively stable (time-invariant) over the course of an observation period (e.g. height, age), while others are time-varying (e.g. workload). Some analyses can incorporate both time-varying and time-invariant variables, while others are limited in this respect – and all analyses that cannot or did not address the multifactorial nature of injury cannot include time-varying and time-invariant variables concurrently. Group difference tests (t-tests, ANOVAs, etc.) may collect time-varying measures, but must aggregate them into a single average for analysis.

Including time-varying and stable variables in the same analysis links closely to between- and within-athlete differences – with the frailty model [52] and multilevel model [62] both used in a way that allowed the researchers to include both. The one exception in our included studies was the generalised estimating equation approach. As mentioned earlier, the GEE assumes an 'overall' effect for each explanatory variable, such that between- and within-athlete differences cannot be differentiated. However, the major benefit to a GEE is that it accounts for the repeated measures for each participant, and can therefore include both time-invariant and time-varying variables for each participant.

Consideration #5 – ILD challenge – handling missing and unbalanced data

Dealing with missing and unbalanced data is a near certainty when dealing with ILD, and is common in applied workload-monitoring settings [67]. The approach to missing data is important from both a methodological and statistical perspective. 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.

A case could be made that group-difference analyses, like the independent samples or paired t-test, also handle unbalanced data across different athletes. For example, Dennis et al (2003) compare the average number of balls bowled per week for injured vs. uninjured bowlers. Some of these bowlers may have contributed 2 weeks of data before being injured, while others contributed a full season of bowling data and were not injured – in this way, there are 'unbalanced' data. However, we contend that these group difference tests aggregate these data across the whole season and force all athletes to only have one observation (usually an aggregated statistic like 'average number of balls bowled per week'. A paired t-test does the same, except it requires all athletes to contribute 2 observations (or in this case, 2 aggregated statistics).

Few authors discussed how they handled missing data in cases of technological difficulties, athlete non-compliance, etc. This is extremely important in cases where authors derive variables like rolling workload averages (e.g. one-week, 'acute', workloads, four-week average, 'chronic', workloads, etc.) [33,36], '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 unreliable. 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 [68].

Challenge #6 – ILD challenge – dependencies created by repeated measures

Collecting ILD in applied sport settings means repeated (often daily) measurements taken from the same athletes – such that observations are clustered within athletes. Comparisons of independent groups, through chi-square tests, independent sample t-tests and one-way/two-way ANOVAs all assume participants contribute a single observation to the analysis and force an aggregated variable (e.g. average number of balls bowled in a week) to conduct the analysis (Dennis, 2003). Similarly, correlation and simple regression (in its linear, logistic, and multinomial forms) assume independence of observations [69]. Paired t-tests and repeated measures ANOVA were used to deal with repeated measurements by comparing the same athletes' workloads at different periods (e.g. the week before injury vs. weeks that did no precede injury).

Of the analyses that addressed this challenge, GEEs were the most common – used in 6 studies. Its ability to handle clustering was also used in one article to control for players clustering within teams [39]. Cox proportional hazard models, used in one article [54], can handle repeated measurements for participants [70]. Multilevel models (Windt et al., 2016) and frailty models – an extension of the Cox- proportional hazards model (Gabbett and Ullah, 2012) – were also used in a single instance each, where repeated measures were clustered within players through a random player effect.

As mentioned in the introduction, there may be additional data dependency created by recurrent injuries [71]. Previous recommendations to handle the recurrent injury challenges have included frailty models [72], and a multistate framework [73]. However, as so few articles reported collecting information on recurrent injuries (n = 5), we focused primarily on the dependencies caused by repeated measures across participants.

Challenge #7 – ILD challenge – incorporating time into the analysis / temporality

One of the most relevant questions in ILD analyses is the way that time is accounted for [1,3]. Some authors used one-way [32] and two-way repeated measures ANOVAs [49] to compare loading in different seasons or season-periods – a very simple way of accounting for time. Repeated measures ANOVAs [46,54,56] and paired t-tests [43,59] also account for time by categorising time-periods as pre-injury blocks or non-injury blocks. Multilevel models have been used to examine change through the interactions of variables with time, but the one multilevel model used in this review did not include time as a covariate [62]. Survival analyses explicitly account for time by calculating the effects of variables on the predicted time-to-event [52,54]. Notably, only one analysis – the frailty model (Gabbett and Ullah, 2012) – adjusted the probability of long-term outcomes (e.g., injury) based on variations after an initial capture of risk, something few traditional analyses accomplish (Cook 2016).

Temporality is also vital in considering potential causal associations. While making causal inferences from observational data is a topic beyond the scope of this paper, temporality is a well-accepted component of causality dating back, at least, to Bradford Hill's 'criteria' [74]. Without temporality – where a postulated cause precedes the outcome – directional associations cannot be made [75,76]. A lack of temporality can also skew associations since it allows for reverse causality. In the workload—injury field, findings that high weekly workloads are sometimes associated with lower odds of injury in a given week [60,62] may be in part because players who get injured in a given week are less likely to accumulate high weekly workloads.

Trying to account for temporality, some researchers have included a latent period – where workload variables are examined for their association with injury occurrence in a given proceeding time window, like the subsequent week [33,62]. While recent work has noted that the length of the latent period may affect model findings [77], it is clear that without some type of latent period, any directional inferences between workloads and injury cannot be made.

Methodological, statistical and reporting considerations

Data aggregation

Data aggregation was common, whether in data preparation, or forced through the analysis. In some cases, researchers aggregated individual level data into team-level measures (total/average workload and injury incidence). Although 32/34 articles collected daily data, most aggregated these daily data into weekly measures, potentially contributing to temporality problems if no latent period was included. Finally, certain analyses (e.g. paired t-tests, simple logistic regression) aggregated data for athletes across an entire year so that workload measures were used to control for exposure [35,61]. Differences in analyses make it impossible to measure the effect of fluctuations in workload and potential impact on injury risk. Further, with no latent period, the directionality of the relationship is unclear. For example, players with high exposure throughout the year were at a lower injury risk than the intermediate group, but it could be interpreted that players who do not sustain an injury throughout the year are more likely to accumulate high total training and match exposures (i.e. higher workload) [35]. Aggregated data may be easier to analyse, but comes at the cost of losing some of the inherent benefits of collecting ILD. As a result, theory-driven questions that relate to daily workload fluctuations and injury risk will become challenging, or impossible to answer.

Checking model assumptions and fit

Authors commonly reported checking was for normality, through Shapiro-Wilk [56] or Kolmogorov-Smirnov tests [32]. Regression modelling was the most common analysis to investigate the workload—injury association. In 8/10 simple logistic regression instances, the authors appear to conduct the analyses using weekly observations without accounting for the dependencies created by repeated-measures across players. In all regression instances, it was uncommon for model assumptions to be checked. Where multiple regression approaches were used, multicollinearity checks were rarely reported – an important consideration since multicollinearity can be a problem when multiple workload variables are simultaneously modelled [41].

Of the papers that modelled data using regression or similar techniques, six described how they assessed model fit. Some authors assessed specificity/sensitivity, or receiver operating characteristics, either on the current data set [61], or future data set (Gabbett, 2010). Other in-sample model fit indices R^2 values [63], Aikeke Information Criteria (AIC) and Bayesian Information Criteria (BIC), which were sometimes mentioned as guiding the model selection process [62].

While many studies may have under-reported how they assessed model assumptions or fit, others [41] provide an example for other researchers to emulate. In fitting a GEE to account for intra-team and intra-player clustering effects, they explained how they selected an appropriate autocorrelation structure, reported how potential quadratic relationships were assessed in the case of non-linear associations, and described checking for potential multicollinearity with defined thresholds (variance inflation factor >10) and for their GEE.

Researcher 'trade-offs', consequences of misalignment

We have used the workload—injury field to highlight seven key themes that emerge from the theoretical model and temporal design (ILD), and shown how some statistical models either cannot, or were not used in a way that addresses these themes. In some cases, misalignment may carry a severe cost – like assumption violations that may bias study results [29]. This is akin to building conclusions on an unstable foundation. Other times, researchers have properly employed their chosen statistical approach, but the approaches themselves are limited, and unable to answer research questions that ILD can address. This is more akin to having a grand building plan and all the necessary supplies, but only using a screwdriver to construct the building.

Regular regression models provide an ideal example of researchers' trade-offs when using traditional statistical analyses on ILD, and the potential costs of misalignment. Although 13 papers used regular regression to analyse the association between workloads and injury outcome, they chose one of three paths when dealing with ILD.

First, many proceeded to analyse each daily or weekly data point as an independent observation – not addressing the violation of the independence assumption (e.g. Hulin et al., 2014, 2016; Rogalski et al., 2013). Second, some researchers aggregated the workload data into an average weekly workload or total workload exposure over the course of the year, such that each participant contributed only one observation to a classic logistic regression [35,61]. Although the regression assumptions aren't violated, workload is aggregated into a single metric, the temporal relationship between workload and injury has been lost, and there is no way to analyse the effects of workload fluctuations on injury risk. Third, some researchers converted individual data to team level data and examined team workloads with team injury incidence in a linear regression [48,63]. In this final case, no differentiation can be made between players or within-players, and inferences are only possible at the team level. This may be sufficient to inform research on the association of workloads and injury at the team level, but the theoretical model underpinning team injury rates may be different than those underpinning individual athletes' injury risk.

Bright spots and future directions

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. Although, 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 full advantage of the benefits of ILD. Functional data analysis [78], machine learning approaches [79,80], generalised estimating equations [81], mixed effects/multilevel/hierarchical models [82], time series analysis [83], survival analyses like shared frailty models [72,84], and time-varying effect models [85] all show promise for ILD.

Since the search for this current review was conducted, there have been some promising developments in the sports medicine field. Recent publications have applied statistical models that more appropriately take advantage of the strengths inherent to ILD, and better align with the theoretical frameworks [79–81,86–88].

Mediation and moderation are causal models which may also contribute to theoretical models [89]. We recently proposed that traditional intrinsic and extrinsic risk factors may act as moderators of the workload—injury association [90]. If that is true, the most appropriate statistical model would be one that incorporates workload measures as the variable of interest, and includes other risk factors as interaction terms that act as 'effect-measure modifiers' [91]. While none of the included articles performed such an analysis, a recent study (not included in this review because it was published after our search) adopted this approach [86]. 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). 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).

We encourage collaboration between applied researchers and statisticians. More sophisticated models may better suit ILD, but come with increased complexity and the need for in-depth understanding that statisticians can provide. However, these complex approaches will only have value where they align with the theoretical models underpinning the specific research field. Based on our included papers, there is an opportunity for improved collaboration, as few researchers referenced any 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 [92].

Review limitations

Previous systematic reviews investigating the workload—injury relationship have documented the challenges of identifying articles through classic systematic review search strategies [7,9]. Heterogeneous key word use and the breadth of sporting contexts have meant previous systematic reviews include many articles post-hoc that were not originally identified by their systematic searches (29 of 67 articles in Jones, 12 of 35 articles in Drew/Finch). Therefore, although we worked to identify articles through 6 systematic reviews and the IOC consensus statement, there may have been potentially eligible articles that were not found.

We used the cut-off for intensive longitudinal data (>20 observations) proposed by Collins (2006). However, there is no universal cut-off for ILD, with previous thresholds of 'more than a handful' [1], ten observations [93], or 40 [3].

It is possible that we are a little unfair to some papers. For example, they 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 (Meeuwisse, 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. Although heuristically appealing, it remains to be seen whether complex systems approaches are effective for analysing workload—injury data. It may also be the case that some environments (e.g. non-professional) lacked the resources to collect multiple variables, limiting their ability to conduct multifactorial analyses.

Finally, it was beyond the scope of this review to list every challenge posed by ILD, and we are not exhaustive in our discussion of different analyses and their capacity to handle the challenges that we did highlight. Where possible, we have tried to identify the themes that are most common within the sports medicine field. Ultimately, the goal of the paper is not that statistical tools should guide our practice, but that statistical tools are more thoughtfully chosen so that the extensive work put into theory building and data collection is not short-changed by a sub-optimal statistical model.

CONCLUSION

We used studies investigating the relationship between workloads and injury as substrate to remind sports medicine researchers how important it is to align their theoretical model, temporal design, and statistical model. In longitudinal research, thoughtfully chosen statistical analyses are those grounded in subject matter theory and that maximise the utility of the collected data. The three most common analyses in our included papers (logistic regression, correlations, and relative risk calculations) addressed one or none of the 3 key theoretical themes, and one or fewer of the 4 inherent challenges of intensive longitudinal data. In this example discipline, researchers have developed sophisticated theories and frequently collect data that will enable them to test these theoretical models. The missing step, and future opportunity for researchers is to avail themselves of all the tools at their disposal – choosing statistical models that address the ILD challenges and answer theory-driven research questions.

Contributions: JW and TG searched for, screened, and identified appropriate systematic reviews and consensus statements. JW identified appropriate original data papers from included systematic reviews/consensus statements. JW and CA performed quality assessment of original data papers and coded the study methods. BZ provided the statistical feedback on the creation of the data extraction spreadsheet and advice on the approach of the methodological review. JW and BZ completed the qualitative assessment of whether authors' use of statistical tools aligned with theoretical or temporal design themes. BS, CA, BZ and KK provided early input during early stages of study development CC contributed to the original idea for the review, and contributed to the discussion section of the manuscript. JW compiled the first draft of the manuscript. CA, CC, BS, BZ, and KK all contributed to critical revision of multiple drafts of the current manuscript.

Acknowledgements: The United States Coalition for the Prevention of Illness & Injury in Sport is a partner in the United States Coalition for the Prevention of Illness and Injury in Sport an International Research Centres for Prevention of Injury and Protection of Athlete Health supported by the International Olympic Committee (IOC).

Technical or equipment support for this study was not provided by any outside companies, manufacturers or organizations.

Funding: JW was a Vanier Scholar funded by the Canadian Institutes of Health Research.

Competing interests: None

Provenance and peer review: Not commissioned, externally peer reviewed.

Ethics approval: No ethics approval was sought for conducting this methodological review.

REFERENCES

1 Walls TA, Schafer JL. *Models for intensive longitudinal data*. Oxford [U.K.]: : Oxford University Press 2006. http://www.loc.gov/catdir/enhancements/fy0636/2005047269-t.html (accessed 13 Apr 2016).

2 Ginexi EM, Riley W, Atienza AA, *et al.* The promise of intensive longitudinal data capture for behavioral health research. *Nicotine Tob Res Off J Soc Res Nicotine Tob* 2014;**16 Suppl 2**:S73–5. doi:10.1093/ntr/ntt273

3 Hamaker EL, Wichers M. No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Curr Dir Psychol Sci* 2017;**26**:10–5. doi:10.1177/0963721416666518

4 Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. *Sports Med* 2014;**44**:139–47. doi:10.1007/s40279-014-0253-z

5 McCall A, Carling C, Nedelec M, *et al.* Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med* 2014;**48**:1352–7. doi:10.1136/bjsports-2014-093439

6 Akenhead R, Nassis GP. Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *Int J Sports Physiol Perform* Published Online First: 9 October 2015. doi:10.1123/ijspp.2015-0331

7 Jones CM, Griffiths PC, Mellalieu SD. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Med* 2016;:1–32. doi:10.1007/s40279-016-0619-5

8 Black GM, Gabbett TJ, Cole MH, *et al.* Monitoring Workload in Throwing-Dominant Sports: A Systematic Review. *Sports Med* 2016;:1–14. doi:10.1007/s40279-016-0529-6

9 Drew MK, Finch CF. The Relationship Between Training Load and Injury, Illness and Soreness: A Systematic and Literature Review. *Sports Med* 2016;:1–23. doi:10.1007/s40279-015-0459-8

10 Gabbett TJ, Whyte DG, Hartwig TB, *et al.* The Relationship Between Workloads, Physical Performance, Injury and Illness in Adolescent Male Football Players. *Sports Med* 2014;**44**:989–1003. doi:10.1007/s40279-014-0179-5

11 Soligard T, Schwellnus M, Alonso J-M, *et al.* How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med* 2016;**50**:1030–41. doi:10.1136/bjsports-2016-096581

12 Keselman HJ, Huberty CJ, Lix LM, *et al.* Statistical Practices of Educational Researchers: An Analysis of Their ANOVA, MANOVA, and ANCOVA Analyses. *Rev Educ Res* 1998;**68**:350–86. doi:10.2307/1170601

13 Weir A, Rabia S, Ardern C. Trusting systematic reviews and meta-analyses: all that glitters is not gold! *Br J Sports Med* 2016;:bjsports-2015-095896. doi:10.1136/bjsports-2015-095896

14 Collins L. Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model. *Annu Rev Psychol* 2006;**57**:505–28. doi:10.1146/annurev.psych.57.102904.190146

15 Lemon SC, Wang ML, Haughton CF, *et al.* Methodological quality of behavioural weight loss studies: a systematic review. *Obes Rev* 2016;:n/a-n/a. doi:10.1111/obr.12412

16 Gabbett TJ, Nassis GP, Oetter E, *et al.* The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data. *Br J Sports Med* 2017;:bjsports-2016-097298. doi:10.1136/bjsports-2016-097298

17 Olivier B, Taljaard T, Burger E, *et al.* Which Extrinsic and Intrinsic Factors are Associated with Non-Contact Injuries in Adult Cricket Fast Bowlers? *Sports Med* 2015;:1–23. doi:10.1007/s40279-015-0383-y

18 Whittaker JL, Small C, Maffey L, *et al.* Risk factors for groin injury in sport: an updated systematic review. *Br J Sports Med* 2015;**49**:803–9. doi:10.1136/bjsports-2014-094287

19 Bahr R, Holme I. Risk factors for sports injuries — a methodological approach. *Br J Sports Med* 2003;**37**:384–92. doi:10.1136/bjsm.37.5.384

20 Bahr R, Krosshaug T. Understanding injury mechanisms: a key component of preventing injuries in sport. *Br J Sports Med* 2005;**39**:324–9. doi:10.1136/bjsm.2005.018341

21 Meeuwisse WH. Assessing Causation in Sport Injury: A Multifactorial Model. *Clin J Sport Med* 1994;**4**:166–170.

22 Meeuwisse WH, Tyreman H, Hagel B, *et al.* A Dynamic Model of Etiology in Sport Injury: The Recursive Nature of Risk and Causation. *Clin J Sport Med* 2007;**17**:215–9. doi:10.1097/JSM.0b013e3180592a48

23 Windt J, Gabbett TJ. How do training and competition workloads relate to injury? The workload—injury aetiology model. *Br J Sports Med* 2016;:bjsports-2016-096040. doi:10.1136/bjsports-2016-096040

24 Bittencourt NFN, Meeuwisse WH, Mendonça LD, *et al.* Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *Br J Sports Med* 2016;:bjsports-2015-095850. doi:10.1136/bjsports-2015-095850

25 Hulme A, Salmon PM, Nielsen RO, *et al.* Closing Pandora's Box: adapting a systems ergonomics methodology for better understanding the ecological complexity underpinning the development and prevention of running-related injury. *Theor Issues Ergon Sci* 2017;**18**:338–59. doi:10.1080/1463922X.2016.1274455

26 Cook C. Predicting future physical injury in sports: it's a complicated dynamic system. *Br J Sports Med* 2016;:bjsports-2016-096445. doi:10.1136/bjsports-2016-096445

27 Sun L, Wu R. Mapping complex traits as a dynamic system. *Phys Life Rev* 2015;**13**:155–85. doi:10.1016/j.plrev.2015.02.007

28 Hoffman L. *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. Routledge 2014. http://www.tandfebooks.com/isbn/9781315744094 (accessed 8 Apr 2016).

29 Wilkinson M, Akenhead R. Violation of statistical assumptions in a recent publication? *Int J Sports Med* 2013;**34**:281. doi:10.1055/s-0032-1331775

30 Anderson L, Triplett-McBride T, Foster C, *et al.* Impact of training patterns on incidence of illness and injury during a women's collegiate basketball season. *J Strength Cond Res* 2003;**17**:734–8.

31 Brooks JHM, Fuller CW, Kemp SPT, *et al.* An assessment of training volume in professional rugby union and its impact on the incidence, severity, and nature of match and training injuries. *J Sports Sci* 2008;**26**:863–73. doi:10.1080/02640410701832209

32 Killen NM, Gabbett TJ, Jenkins DG. Training Loads and Incidence of Injury During the Preseason in Professional Rugby League Players: *J Strength Cond Res* 2010;**24**:2079–84. doi:10.1519/JSC.0b013e3181ddafff

33 Hulin BT, Gabbett TJ, Blanch P, *et al.* Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *Br J Sports Med* 2014;**48**:708–12. doi:10.1136/bjsports-2013-092524

34 Fuller CW, Ekstrand J, Junge A, *et al.* Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med* 2006;**40**:193–201. doi:10.1136/bjsm.2005.025270

35 Arnason A, Sigurdsson SB, Gudmundsson A, *et al.* Risk Factors for Injuries in Football. *Am J Sports Med* 2004;**32**:5S–16S. doi:10.1177/0363546503258912

36 Bowen L, Gross AS, Gimpel M, *et al.* Accumulated workloads and the acute:chronic workload ratio relate to injury risk in elite youth football players. *Br J Sports Med* 2016;:bjsports-2015-095820. doi:10.1136/bjsports-2015-095820

37 Bresciani G, Cuevas MJ, Garatachea N, *et al.* Monitoring biological and psychological measures throughout an entire season in male handball players. *Eur J Sport Sci* 2010;**10**:377–84. doi:10.1080/17461391003699070

38 Brink MS, Nederhof E, Visscher C, *et al.* Monitoring load, recovery, and performance in young elite soccer players. *J Strength Cond Res Natl Strength Cond Assoc* 2010;**24**:597–603. doi:10.1519/JSC.0b013e3181c4d38b

39 Clausen MB, Zebis MK, Møller M, *et al.* High Injury Incidence in Adolescent Female Soccer. *Am J Sports Med* 2014;**42**:2487–94. doi:10.1177/0363546514541224

40 Colby MJ, Dawson B, Heasman J, *et al.* Accelerometer and GPS-Derived Running Loads and Injury Risk in Elite Australian Footballers: *J Strength Cond Res* 2014;**28**:2244–52. doi:10.1519/JSC.0000000000000362

41 Cross MJ, Williams S, Trewartha G, *et al.* The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *Int J Sports Physiol Perform* Published Online First: 26 August 2015. doi:10.1123/ijspp.2015-0187

42 Dennis R, Farhart P, Clements M, *et al.* The relationship between fast bowling workload and injury in first-class cricketers: a pilot study. *J Sci Med Sport Sports Med Aust* 2004;**7**:232–6.

43 Dennis R, Farhart R, Goumas C, *et al.* Bowling workload and the risk of injury in elite cricket fast bowlers. *J Sci Med Sport* 2003;**6**:359–67. doi:10.1016/S1440-2440(03)80031-2

44 Dennis RJ, Finch CF, Farhart PJ. Is bowling workload a risk factor for injury to Australian junior cricket fast bowlers? *Br J Sports Med* 2005;**39**:843–6. doi:10.1136/bjsm.2005.018515

45 Duhig S, Shield AJ, Opar D, *et al.* Effect of high-speed running on hamstring strain injury risk. *Br J Sports Med* 2016;:bjsports-2015-095679. doi:10.1136/bjsports-2015-095679

46 Ehrmann FE, Duncan CS, Sindhusake D, *et al.* GPS and Injury Prevention in Professional Soccer. *J Strength Cond Res* Published Online First: 2015. doi:10.1519/JSC.0000000000001093

47 Gabbett TJ. The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *J Strength Cond Res* 2010;**24**:2593–603. doi:10.1519/JSC.0b013e3181f19da4

48 Gabbett TJ. Influence of training and match intensity on injuries in rugby league. *J Sports Sci* 2004;**22**:409–17. doi:10.1080/02640410310001641638

49 Gabbett TJ. Reductions in pre-season training loads reduce training injury rates in rugby league players. *Br J Sports Med* 2004;**38**:743–9. doi:10.1136/bjsm.2003.008391

50 Gabbett TJ, Jenkins DG. Relationship between training load and injury in professional rugby league players. *J Sci Med Sport* 2011;**14**:204–9. doi:10.1016/j.jsams.2010.12.002

51 Gabbett TJ, Domrow N. Relationships between training load, injury, and fitness in sub-elite collision sport athletes. *J Sports Sci* 2007;**25**:1507–19. doi:10.1080/02640410701215066

52 Gabbett TJ, Ullah S. Relationship Between Running Loads and Soft-Tissue Injury in Elite Team Sport Athletes. *J Strength Cond Res* 2012;**26**:953–60. doi:10.1519/JSC.0b013e3182302023

53 Hulin BT, Gabbett TJ, Caputi P, *et al.* Low chronic workload and the acute:chronic workload ratio are more predictive of injury than between-match recovery time: a two-season prospective cohort study in elite rugby league players. *Br J Sports Med* 2016;:bjsports-2015-095364. doi:10.1136/bjsports-2015-095364

54 Malisoux L, Frisch A, Urhausen A, *et al.* Monitoring of sport participation and injury risk in young athletes. *J Sci Med Sport* 2013;**16**:504–8. doi:10.1016/j.jsams.2013.01.008

55 Mallo J, Dellal A. Injury risk in professional football players with special reference to the playing position and training periodization. *J Sports Med Phys Fitness* 2012;**52**:631–8.

56 Murray NB, Gabbett TJ, Townshend AD, *et al.* Individual and combined effects of acute and chronic running loads on injury risk in elite Australian footballers. *Scand J Med Sci Sports* 2016;:n/a-n/a. doi:10.1111/sms.12719

57 Murray NB, Gabbett TJ, Townshend AD. Relationship Between Pre-Season Training Load and In-Season Availability in Elite Australian Football Players. *Int J Sports Physiol Perform* 2016;:1–21. doi:10.1123/ijspp.2015-0806

58 Rogalski B, Dawson B, Heasman J, *et al.* Training and game loads and injury risk in elite Australian footballers. *J Sci Med Sport* 2013;**16**:499–503. doi:10.1016/j.jsams.2012.12.004

59 Saw R, Dennis RJ, Bentley D, *et al.* Throwing workload and injury risk in elite cricketers. *Br J Sports Med* 2011;**45**:805–8. doi:10.1136/bjsm.2009.061309

60 Veugelers KR, Young WB, Fahrner B, *et al.* Different methods of training load quantification and their relationship to injury and illness in elite Australian football. *J Sci Med Sport* 2015;**19**:24–8. doi:10.1016/j.jsams.2015.01.001

61 Visnes H, Bahr R. Training volume and body composition as risk factors for developing jumper's knee among young elite volleyball players. *Scand J Med Sci Sports* 2013;**23**:607–13. doi:10.1111/j.1600-0838.2011.01430.x

62 Windt J, Gabbett TJ, Ferris D, *et al.* Training load--injury paradox: is greater preseason participation associated with lower in-season injury risk in elite rugby league players? *Br J Sports Med* 2016;:bjsports-2016-095973. doi:10.1136/bjsports-2016-095973

63 Owen AL, Forsyth JJ, Wong DP, *et al.* Heart rate-based training intensity and its impact on injury incidence among elite-level professional soccer players. *J Strength Cond Res Natl Strength Cond Assoc* 2015;**29**:1705–12. doi:10.1519/JSC.0000000000000810

64 Ghisletta P, Spini D. An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. *J Educ Behav Stat* 2004;**29**:421–37. doi:10.3102/10769986029004421

65 Bellera CA, MacGrogan G, Debled M, *et al.* Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol* 2010;**10**:20. doi:10.1186/1471-2288-10-20

66 Hulme A, Finch CF. From monocausality to systems thinking: a complementary and alternative conceptual approach for better understanding the development and prevention of sports injury. *Inj Epidemiol* 2015;**2**. doi:10.1186/s40621-015-0064-1

67 Buchheit M. Applying the acute:chronic workload ratio in elite football: worth the effort? *Br J Sports Med* 2016;:bjsports-2016-097017. doi:10.1136/bjsports-2016-097017

68 Collins N, Crossley K, Beller E, *et al.* Foot orthoses and physiotherapy in the treatment of patellofemoral pain syndrome: randomised clinical trial. *Br J Sports Med* 2009;**43**:163–8. doi:10.1136/bmj.a1735

69 McDonald JH. *Handbook of biological statistics*. Sparky House Publishing Baltimore, MD 2009. http://biostathandbook.com/HandbookBioStatSecond.pdf

70 Therneau TM, Grambsch PM. *Modeling survival data: extending the Cox model*. New York: : Springer 2000.

71 Finch CF, Cook J. Categorising sports injuries in epidemiological studies: the subsequent injury categorisation (SIC) model to address multiple, recurrent and exacerbation of injuries. *Br J Sports Med* 2014;**48**:1276–80. doi:10.1136/bjsports-2012-091729

72 Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med* 2014;**48**:1287–93. doi:10.1136/bjsports-2011-090803

73 Shrier I, Steele RJ, Zhao M, *et al.* A multistate framework for the analysis of subsequent injury in sport (M-FASIS). *Scand J Med Sci Sports* 2016;**26**:128–39. doi:10.1111/sms.12493

74 Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965;**58**:295–300.

75 Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg Themes Epidemiol* 2005;**2**:11. doi:10.1186/1742-7622-2-11

76 Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;**95 Suppl 1**:S144-150. doi:10.2105/AJPH.2004.059204

77 Carey DL, Blanch P, Ong K-L, *et al.* Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk. *Br J Sports Med* 2016;:bjsports-2016-096309. doi:10.1136/bjsports-2016-096309

78 Trail JB, Collins LM, Rivera DE, *et al.* Functional data analysis for dynamical system identification of behavioral processes. *Psychol Methods* 2014;**19**:175–87. doi:10.1037/a0034035

79 Carey DL, Ong K-L, Whiteley R, *et al.* Predictive modelling of training loads and injury in Australian football. *ArXiv170604336 Stat* Published Online First: 14 June 2017.http://arxiv.org/abs/1706.04336

80 Rossi A, Pappalardo L, Cintia P, *et al.* Effective injury prediction in professional soccer with GPS data and machine learning. *ArXiv170508079 Stat* Published Online First: 23 May 2017.http://arxiv.org/abs/1705.08079

81 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. *J Sci Med Sport* Published Online First: 2017.http://www.sciencedirect.com/science/article/pii/S144024401730436X

82 Smith DM, Walls TA. Multiple Time Scale Models in Sport and Exercise Science. *Meas Phys Educ Exerc Sci* 2016;**20**:185–99. doi:10.1080/1091367X.2016.1195390

83 de Haan-Rietdijk S, Kuppens P, Hamaker EL. What's in a Day? A Guide to Decomposing the Variance in Intensive Longitudinal Data. *Front Psychol* 2016;**7**. doi:10.3389/fpsyg.2016.00891

84 Nielsen RØ, Malisoux L, Møller M, *et al.* Shedding Light on the Etiology of Sports Injuries: A Look Behind the Scenes of Time-to-Event Analyses. *J Orthop Sports Phys Ther* 2016;**46**:300–11. doi:10.2519/jospt.2016.6510

85 Biddle SJH, Edwardson CL, Wilmot EG, *et al.* A Randomised Controlled Trial to Reduce Sedentary Time in Young Adults at Risk of Type 2 Diabetes Mellitus: Project STAND (Sedentary Time ANd Diabetes). *PLOS ONE* 2015;**10**:e0143398. doi:10.1371/journal.pone.0143398

86 Møller M, Nielsen RO, Attermann J, *et al.* Handball load and shoulder injury rate: a 31-week cohort study of 679 elite youth handball players. *Br J Sports Med* 2017;**51**:231–7. doi:10.1136/bjsports-2016-096927

87 Stares J, Dawson B, Peeling P, *et al.* Identifying high risk loading conditions for in-season injury in elite Australian football players. *J Sci Med Sport* Published Online First: 25 May 2017. doi:10.1016/j.jsams.2017.05.012

88 Williams S, Trewartha G, Cross MJ, *et al.* Monitoring What Matters: A Systematic Process for Selecting Training Load Measures. *Int J Sports Physiol Perform* 2016;:1–20. doi:10.1123/ijspp.2016-0337

89 Wu AD, Zumbo BD. Understanding and Using Mediators and Moderators. *Soc Indic Res* 2007;**87**:367. doi:10.1007/s11205-007-9143-1

90 Windt J, Zumbo BD, Sporer B, *et al.* Why do workload spikes cause injuries, and which athletes are at higher risk? Mediators and moderators in workload–injury investigations. *Br J Sports Med* 2017;:bjsports-2016-097255. doi:10.1136/bjsports-2016-097255

91 Nielsen RO, Bertelsen ML, Møller M, *et al.* Training load and structure-specific load: applications for sport injury causality and data analyses. *Br J Sports Med* 2017;:bjsports-2017-097838. doi:10.1136/bjsports-2017-097838

92 Casals M, Finch CF. Sports Biostatistician: a critical member of all sports science and medicine teams for injury prevention. *Inj Prev* 2016;:injuryprev-2016-042211. doi:10.1136/injuryprev-2016-042211

93 Tan X, Shiyko MP, Li R, *et al.* A time-varying effect model for intensive longitudinal data. *Psychol Methods* 2012;**17**:61–77. doi:10.1037/a0025814

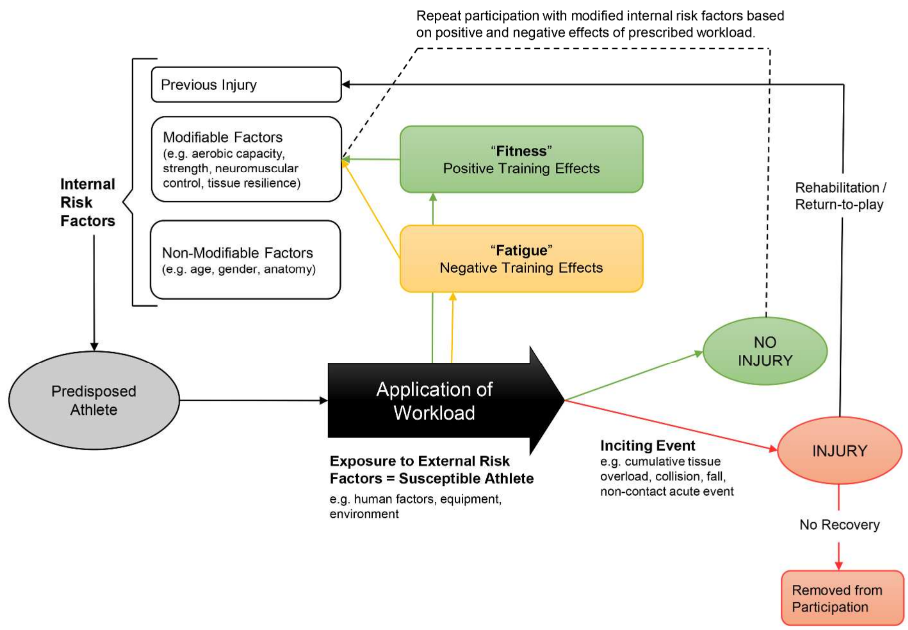

Figure 1 – The workload—injury aetiology model. Key features include the multifactorial nature of injury, between- and within- athlete differences in risk, and a recursive loop.

299x199mm (300 x 300 DPI)

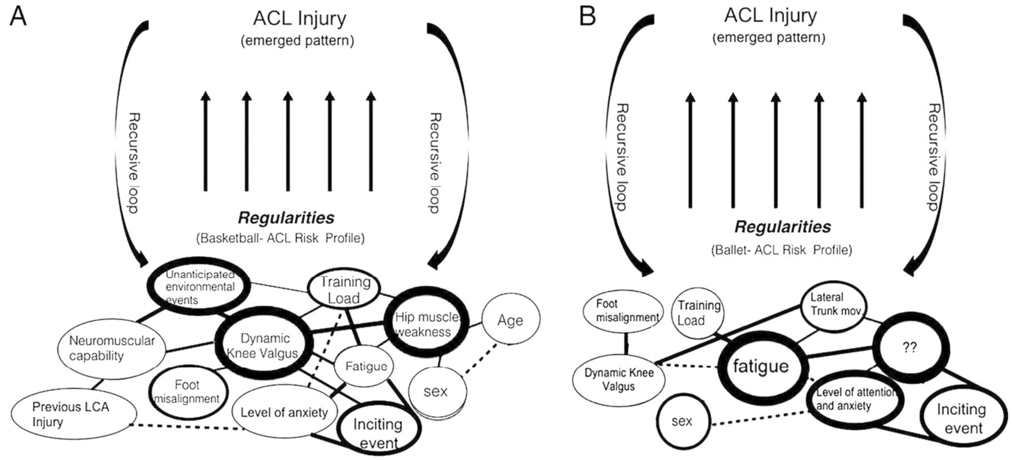

Figure 2 – Complex systems model of athletic injury. Web of determinants are shown for an ACL injury in basketball players (A), and in a ballet dancer (B)

152x69mm (300 x 300 DPI)

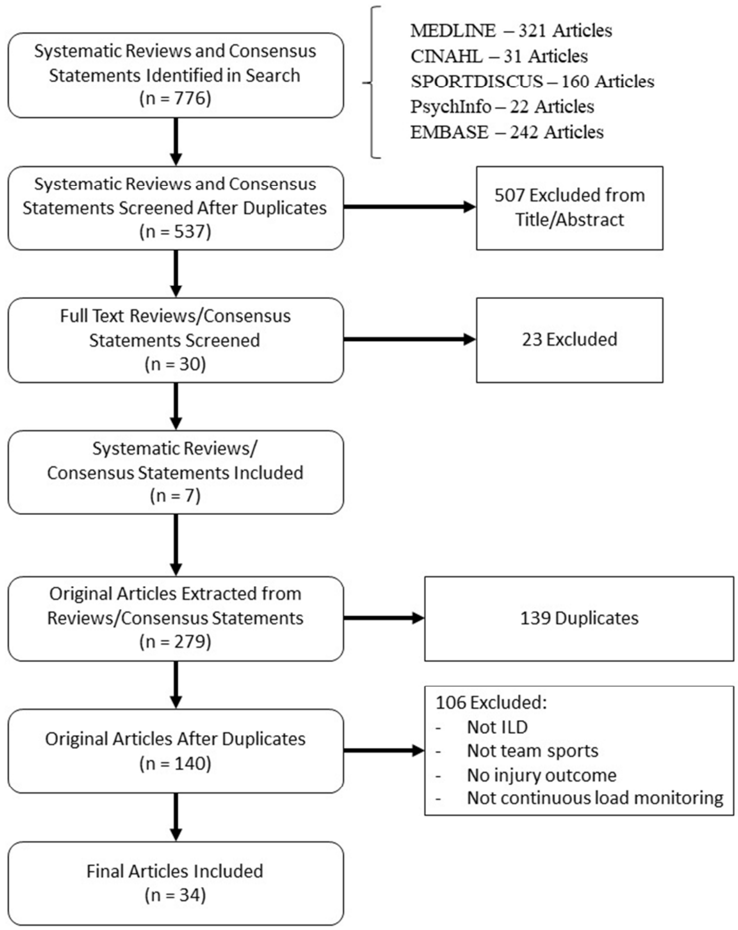

Figure A1 – Flow chart of included articles

80x96mm (300 x 300 DPI)

APPENDIX 1 – Systematic Search Strategy and Article Selection

Table A1 – Example search categories and terms:

Category	Search Terms
Population	Athlet*, sport*
Method	Systematic review, Consensus statement
Outcome (Injury)	Injur*, illness*, strain, sprain, incidence, overuse, overreach*, accidents, stress, wellness, recover*
Workload	Training, resistance training, external load, internal load, workload, acute:chronic workload ratio, congested calendar, physical exertion, session RPE, global position systems, accelerometry, intensity, duration, physical fitness, fatigue

--- INSERT FIGURE A1 HERE ---

Figure A1 – Flow chart of included articles

APPENDIX 2 – TABLE A2: Expanded Table of 3-Fold Alignment (With Descriptions)

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Correlation (Pearson and Spearman)	10	X	X	X	X	X	X	X	7 of the 10 articles correlated team loads with injury incidence (team-level). Of those at the individual level, 2 looked at # of pre-season sessions and % in-season completed, and 1 looked at training load and injury subscale on the REST-Q.
Anderson - 2003 Bresciani - 2010 Books - 2008 Gabbett - 2004 Gabbett Jenkins - 2011 Killen 2010 Mallo Dellal - 2012 Murray Gabbett - 2016 Owen - 2015 Windt – 2016		Can only handle one x and one y variable.	Correlation could be at the team level (training load and # of injuries), or on individual level with quantitative outcome, but cannot differentiate within/between athlete differences.	No interactions between multiple predictors	No, assumes independent observations.	Assumes one observation per research participant/study unit, so participants couldn't have different numbers of observations.	In this case the correlation assumes independent observations so dependency is not taken into account.	Could, through having time as one of the variables, but cannot account for temporality	
Unpaired t-test	6	X	X	X	X	X	X	X	Generally, compared injured and uninjured players across a season. A group comparison of loading across the year fails to account for between/within athlete considerations and doesn't specify temporality.
Dennis et al., 2003 Dennis et al., 2005 Duhig et al - 2016 Owen et al., 2015 Saw et al., 2011 Visnes Bahr., 2013		Can only handle one x (grouping) and one y (outcome) variable	Only between-athlete differences (injured vs. uninjured)	No interactions between multiple predictors	Independent samples t-test compare group means on either a time-varying (e.g average workload) or time-invariant variable (e.g. height), not both.	No, assumes one observation per research participant/study unit, so by design forces a balanced set (1 observation per participant)	By definition, assumes independence	None included time in the analysis.	

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Chi-Square Tests	1	X	X	X	X	X	X	X	
Murray - 2016 - IJSPP		Only examines load groups and injury incidence	Examines differences in injury incidence across different load groups only	No interactions between multiple predictors	Included load groups and injury incidence only	Forces 1 observation (aggregated variable) per participant	Designed for independent observations and groups	Only included load groups and injury incidence	
Relative Risk Calculations	8	O	X	X	X	X	X	X	Many of these RR approaches seem to use RRs that traditionally require independence, but do not account for this in their analysis.
Bowen - 2016 Dennis - 2003 Dennis - 2005 Hulin - 2014 Hulin - 2016 Hulin - 2016 Murray - 2016 - Scand.		In some cases, authors examined relative risks of loading groups after subdividing across another variable, like chronic workload, making it multifactorial. Other authors only examined risks across load groups.	No differentiation, and independence assumed	No interactions between multiple predictors	Only loading (time-varying variables included)	Assumes independence of observations	Assumes independence	Uses weeks as unit of analysis but no incorporation of time into the calculations	

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Regression (logistic, linear, multinomial)	13	O	X	X	X	X	X	X	
Arnason 2004 - Regular logistic Bowen - 2016 - Regular logistic Brink - 2010 - Regular multinomial Colby - 2014 - Regular logistic Duhig - 2016 - Regular logistic Gabbett Domrow - 2007 - Regular linear Hulin - 2014 - Regular logistic Hulin - 2016 - Regular logistic Hulin - 2016 - Regular logistic Murray - 2016 - Regular logistic (Scand) Owen - 2015 - Regular linear Rogalski - 2013 - Regular logistic Visnes Bahr - 2013 - Regular logistic		Some authors include multiple variables, others only single load measurements independently	Assume independent observations, so cannot examine within-athlete differences	None included interactions between predictors	Assumes 1 observation per unit Becomes time invariant on aggregation In this case, some authors have included both, but in doing so violate the independence assumption	No, assumes one observation per athlete or research unit	Assumes independence. Visnes & Bahr (2013) do take it into account through logistic regression, but do not have the benefit of ILD	Assumes 1 observation per unit	
Paired t-test	2	X	X	X	X	X	√	√	
Saw et al., 2011 Dennis et al., 2003		No, only one variable (before/after) and outcome (load)	Only examines within-athlete differences	No interactions between multiple predictors	No - only time-varying variables	All subjects must have 2 'observations'	By definition, accounts for repeated measures through paired sample	Time is incorporated as pre-injury and 'injury' blocks of time	

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Repeated measures ANOVA (One- or two-way ANOVA allowing for between- and within)	5	O	O	X	O	X	√	√	Killen used 1-way ANOVA to compare the load in early pre-season compared to late pre-season, and used chi-square analyses to compare injury rate in the early and late pre-season. Uses the two separate analyses to tentatively link load leads to injury. Gabbett, 2004 performed a 2-way ANOVA comparing loads (season X month), so time is included.
Ehrman - 2016 Gabbett - 2004 Malisoux - 2013 Murray - 2016 Killen - 2011		Murray (2016) compared part of season by training load group, Malisoux (2013) and Ehrmann (2016) only compared injury and pre-injury blocks	In some cases (Murray, 2016), a two-way repeated measures ANOVA can examine between and within athlete differences in risk	No interactions between multiple predictors	Murray (2013) compared load group by season period	Assume sphericity	Yes, by definition	Yes, as season period, or as pre-injury and injury period	
Cox proportional hazards model	1	√	X	X	X	√	√	√	
Malisoux - 2013		Included volume and intensity of training along with age and sex	Cox PH conducted at the team/school level examined between-athlete differences	No interactions between factors	Only included average weekly load and average intensity	Can handle unbalanced data	By using time-to-event as the outcome, Cox-PH robust to this dependency	Uses time to event in analysis	

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				**Statistical summary and typical uses**
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Generalised estimating equations (modeled through logistic and poisson regression)	6	O	X	X	O	√	√	O	Most GEE approaches accounted for repeated measures, but Clausen et al (2014) used them to cluster players within teams), averaging exposure throughout the entire season.
Clausen - 2014 Cross - 2016 Dennis - 2004 Gabbett - 2010 Gabbett Domrow - 2007 Veugeleurs - 2016		Most authors used multiple variables with GEEs, although some only used GEE to account for repeated measures (Dennis, 2004, Gabbett Domrow, 2007)	Provides an 'average' effect for all athletes, but controls for the clustering	No complex interactions between factors	Some authors only predicted injury (y/n) based on workload variables	GEEs handle unbalanced data well	GEE accounts for clustering	Do not incorporate time explicitly into the modelling process, and none of the authors included time in the model.	
Multilevel Modeling	1	√	√	X	√	√	√	X	
Windt et al - 2016		Multiple physical outputs included and pre-season training	Within-athlete risks determined from level 1 variables, between-athlete from level 2	No complex interactions between factors	Pre-season and player variables along with training variables	Multilevel models are robust to missing/unbalanced data	Via random effects for each player	Analysed weekly risk and subsequent week risk; did not directly incorporate time	
Frailty model	1	√	√	X	√	√	√	√	Included a many physical output metrics that were dichotomised into high/low categories and calculated RRs. Multicollinearity was not considered
Gabbett Ullah - 2012		Yes, previous injury and physical outputs	Unclear (I think yes due to the frailty term)	No complex interactions between factors	Yes (injury history)	Yes (robust to unbalanced data)	Yes (frailty model allows for dependency of recurrent events)	Yes, since this is a time-to-event analysis.	

BMJ Open

Getting the most out of intensive longitudinal data: A methodological review of workload—injury studies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022626.R1
Article Type:	Research
Date Submitted by the Author:	24-Jul-2018
Complete List of Authors:	Windt, Johann; University of British Columbia, Experimental Medicine Ardern, CL; Linkoping University, Division of Physiotherapy Gabbett, Tim; Gabbett Performance Solutions; University of Southern Queensland, Institute for Resilient Regions Khan, Karim; The University of British Columbia, Centre for Hip Health & Mobility Cook, Chad; Duke University, Department of Orthopedics Sporer, Ben; University of British Columbia, Family Practice Zumbo, Bruno; University of British Columbia, Measurement, Evaluation, and Research Methodology Program
Primary Subject Heading:	Sports and exercise medicine
Secondary Subject Heading:	Research methods
Keywords:	STATISTICS & RESEARCH METHODS, PREVENTIVE MEDICINE, SPORTS MEDICINE

Running Title: ALIGNING THEORY, DATA COLLECTION, AND STATISTICAL ANALYSES IN SPORTS MEDICINE RESEARCH USING LONGITUDINAL DATA

Getting the most out of intensive longitudinal data: A methodological review of workload—injury studies

Johann Windt,[1,2,3] Clare L Ardern,[4,5] Tim J. Gabbett,[6,7] Karim M. Khan,[1,8] Chad E. Cook,[9] Ben Sporer,[8,10] Bruno D. Zumbo[11]

[1] Experimental Medicine Program, University of British Columbia, Vancouver, Canada

[2] United States Olympic Committee, Colorado Springs, United States of America

[3] United States Coalition for the Prevention of Illness and Injury in Sport, Colorado Springs, United States of America

[4] Division of Physiotherapy, Linköping University, Linköping, Sweden

[5] School of Allied Health, La Trobe University, Melbourne, Australia

[6] Gabbett Performance Solutions, Brisbane, Australia

[7] University of Southern Queensland, Institute for Resilient Regions, Ipswich, Australia

[8] Department of Family Practice, University of British Columbia

[9] Department of Orthopaedics, Duke University, Durham, United States of America [10] Vancouver Whitecaps Football Club, Vancouver, British Columbia

[11] Measurement, Evaluation, and Research Methodology Program, University of British Columbia, Vancouver, Canada

Corresponding Author:

Johann Windt, MSc, CSCS (ORCID ID: 0000-0001-5517-2609)

1 Olympic Plaza, Colorado Springs, CO 80909

Email: johannwindt@gmail.com

Key Words: Methodology, workloads, training load, athletic injury, statistics

Word Count: *8323*

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.

Results: Statistical methods such as correlations, t-tests, and simple linear/logistic regression were commonly used. However, these methods did not adequately address the (1) themes of theoretical models underlying workloads and injury, nor the (2) temporal design challenges (ILD). Although time-to-event analyses (e.g. Cox Proportional Hazards and frailty models) and multilevel modelling are better-suited for ILD, these were utilized in fewer than a 10% of the studies (n=3).

Conclusions: Rapidly accelerating availability of intensive longitudinal data is the norm in many fields of health care delivery and thus health research. These data present an opportunity to better address research questions, especially when appropriate statistical analyses are chosen.

STRENGTHS AND LIMITATIONS

- As intensive longitudinal data become increasingly common across disciplines, catalysed by technological advances, this methodological review provides researchers with several considerations when determining how to analyse these data.
- Whereas systematic reviews provide a quantitative synthesis of research findings, they do not account for the statistical approaches used in the original studies. Therefore, methodological reviews like this one fill an important void in the literature to highlight key shortcomings and ways forward from a methodological and statistical perspective.
- By choosing a homogenous group of papers – prospective cohort studies in team sports that collected intensive longitudinal data – we were able to focus more directly on the statistical analyses that authors employed.
- It was beyond the scope of this review to list every challenge posed by intensive longitudinal data, and we are not exhaustive in our discussion of different analyses and their capacity to handle the challenges that we did highlight.

INTRODUCTION

Intensive longitudinal data (ILD) are being collected more frequently in various research areas [1], catalysed by technological advancements that simplify data collection and analysis [2]. By collecting data repeatedly on the same participants, researchers are enabled to answer more detailed research questions, particularly regarding phenomena that change or fluctuate over time. However, arriving at these answers requires researchers to overcome the challenges of analysing ILD, which include: (1) the dependencies created by repeated measures, (2) missing/unbalanced data, (3) separating between- and within-person effects, (4) time-varying and time-invariant (stable) factors, and (5) specifying the role of time/temporality [3].

The field of exercise and sports medicine provides one specific example which can illustrate principles that apply to the use of intensive longitudinal data broadly. In the field of sports performance, technological advances mean that a plethora of physiological, psychological and physical data are conveniently available from athletes [4,5]. As one example, of 48 professional football clubs that responded to a survey on player monitoring, 100% reported collecting daily global positioning system (GPS) and heart rate (HR) data [6].

One research question that has gained a great deal of interest in the last decade is how athletes' training and competition workloads relate to injury risk. Since athletes' training and injury risk continually varies over time, many researchers have used prospective cohort studies to collect and analyse ILD to answer this question [7]. There is moderate evidence from systematic reviews and an International Olympic Committee (IOC) consensus statement suggesting a positive relationship between injury rates and high training workloads, increased risk of injury with low workloads, and a pronounced increase in injury risk associated with rapid workload increases [7–11]. However, such systematic reviews do not consider the statistical approaches used in included studies [12]. Choosing the wrong statistical analysis or poorly implementing an otherwise correct one (e.g. violating statistical assumptions) can bias results and create false conclusions. Even a perfectly performed systematic review cannot compensate for poorly designed, or poorly analysed studies [13].

Longitudinal data analysis is most effective when the chosen statistical approach aligns with the frequency of data collection and with the theoretical model underpinning the research question (See Box 1) [14]. Therefore, we used this lens to evaluate whether the statistical models employed in prospective cohort studies using ILD to investigate the relation between athletic workloads and injury were optimal. We had 3 aims: (1) to summarise researchers' data collection, methodological, statistical, and reporting practices [12,15]; (2) to evaluate the degree to which the adopted statistical analyses fit within Collins' three-fold alignment [see Box 1]; and (3) to provide recommendations for future investigations in the field.

Box 1: Theoretical model – temporal design – statistical model.

In a landmark, highly-cited paper, Professor Linda Collins described how aligning the (1) theoretical model (subject matter theory), (2) temporal design (data collection strategy/timing), and (3) statistical model (analytical strategy) is crucial when analysing longitudinal data [14]. *For example, if researchers (1) theorize that a given physiological variable fluctuates every hour, (2) data must be collected at least on an hourly basis. If researchers measure participants once a day, they will miss virtually all the hourly fluctuations that their theories predict. Once researchers have collected their hourly data, they should (3) select a statistical strategy that enables them to examine the relationship between these fluctuations and the outcome of interest. As Collins noted, perfect alignment of these 3 components may not be possible, but it provides researchers a target, and readers a lens through which longitudinal research can be evaluated.*

METHODS

Article selection

We systematically searched the literature (MEDLINE, CINAHL, SPORTDISCUS, PsychInfo, and EMBASE) (December 10, 2016) to identify systematic reviews and consensus statements that investigated the relationship between workloads and athletic injuries, with the aim of extracting all original articles included in these reviews that met our inclusion criteria. A summary of the systematic search and article selection process is described in Appendix 1 (Table A1 and Figure A1), and the full systematic search is available from the authors.

A priori, we operationally defined 'workload' as either external – the amount of work completed by the athlete (e.g. distance run, hours completed, etc.), or internal – the athlete's response to a given external workload (e.g. session rating of perceived exertion, heart-rate based measures, etc.). We acknowledge that athlete self-reported measures often evaluate how athletes are handling training demands and may be referred to as 'internal' load measures, but we considered these perceptual wellbeing measures as a distinct step from quantifying athletes' internal or external workloads [16]. Athletic injuries have been diversely defined in the literature, so we operationally defined athletic injury as any article that reported measuring 'injury', regardless of their specific definition (e.g. time loss, medical attention, etc).

Two authors (JW + TG) screened the titles/abstracts of the systematic reviews. Where necessary, the full texts were retrieved to determine whether they should be included. A total of 6 systematic reviews [7–10,17,18] and 1 consensus statement [11] were identified that included at least 1 article meeting the inclusion criteria.

We extracted and reviewed the full texts of all the original studies included (n=279) in these 7 papers. For our analysis, we included all the original articles that met the following criteria:

1) Original articles were prospective cohort studies that examined the relationship between at least 1 measure of internal or external workload (as defined above) and athletic injury. Since theoretical models describe the recursive nature of injury risk with each training or competition exposure, workloads had to be continually monitored and include both training and match workloads for the same athletes. Although some athletes may have entered or left the group during the study period (e.g. through retirement or trades to other teams) the same team/group of athletes had to be followed throughout the study period, as opposed to repeated cross-sectional snapshots of different cohorts.
2) Articles collected intensive longitudinal data. We defined intensive longitudinal data as >20 observations per athlete [14].
3) Articles studied team sport athletes. We chose team sports because (1) there are high amounts of ILD collected in applied team sport settings [6], and (2) the majority of workload—injury studies are in team

sport athletes (Jones, 2016). Military populations and individual sports (e.g. distance running) were excluded due to the differences in task requirements and operating environment.

Patient and public involvement

As a methodological review, there was no patient or public involvement in this current investigation.

Article coding and description

To describe the methodological, statistical, and reporting approaches utilized in each article, two authors (JW + CA) reviewed all the included papers and extracted 50 items of information for each article. These items included publication year, journal, variable operationalization (e.g. internal vs. external load measures, injury definition, etc), methodological approaches, statistical analyses implemented, reported findings, and more. To ensure consistency between coders, 10 articles were randomly selected, coded independently by both reviewers, and compared to assess agreement. Discrepancies were discussed by the two coders and an additional 5 articles were randomly selected and coded independently. The remaining articles were coded by JW and checked by CA.

Assessing how statistical models aligned with Collins' threefold framework

To evaluate the statistical approaches used in this field, we first identified the key themes and challenges within the theoretical models and temporal design features within the workload—injury field, then developed a qualitative assessment to evaluate the statistical approaches.

Collins Component 1: The theoretical models that underpin athletic workloads and injury risk (in brief)

Briefly, we identified at least 3 key elements of athletic injury aetiology models. First, *sports injuries are multifactorial* [19–21]. Aetiology models since 1994 have all explained between-athlete differences in injury risk by identifying a host of 'internal' (e.g. athlete characteristics, psychological wellbeing, previous injury) and 'external' (e.g. opponent behaviour, playing surface) risk factors. More recently, Meeuwisse et al.'s dynamic recursive model [22] and the workload—injury aetiology model [23], have highlighted the recurrent nature of injury risk, meaning each athlete's injury risk (i.e. within-athlete risk) also fluctuates continually as they train or compete in their sport (Figure 1). Thus, a second theme is that *injury risk differs between- and within-athletes*. Finally, more recent injury aetiology models have highlighted *injury risk as a complex, dynamic system* (Figure 2)

[24,25]. Complex systems, as in weather forecasting or biological systems [26,27], possess many key features, including an open-system, inherent non-linearity between variables and outcomes, recursive loops where the system output becomes the new system input, self-organization where regular patterns (risk profiles) may emerge for given outcomes (emergent pattern), and uncertainty [24].

--- INSERT FIGURE 1 HERE ---

Figure 1 – The workload—injury aetiology model. Key features include the multifactorial nature of injury, between- and within- athlete differences in risk, and a recursive loop.

--- INSERT FIGURE 2 HERE ---

Figure 2 – Complex systems model of athletic injury. Web of determinants are shown for an ACL injury in basketball players (A), and in a ballet dancer (B)

Collins' 2nd Component – Temporal design / data collection

The theoretical models relating workloads and injury illustrate a continuously fluctuating injury risk, with many variables that influence risk on a daily or weekly basis [22–24]. Thus, if researchers want to investigate the association between workloads and injuries, these data must be collected frequently enough to observe changes in these variables as they occur (temporal design). With technological advances, athletes' physiological, psychological and physical variables are now often collected on a daily, weekly or monthly basis, along with ongoing injury surveillance data [4,5]. Therefore, in the workload—injury field, the theoretical models (injury aetiology models that describe regular fluctuation in workloads and injury risk) and the temporal design (frequent, often daily, data collection) are often well-aligned, especially in prospective cohort studies using ILD. This leaves us to consider only whether Professor Collins' third component – the statistical model – aligns with these first two.

Collins' Component 3 – Statistical model

From the theoretical aetiology models underpinning the workload—injury association, we highlighted three key themes to consider when choosing a statistical model: (1) injury risk is multifactorial, (2) between-athlete and within-athlete differences in injury risk fluctuate regularly, and (3) injury risk may be considered a complex, dynamic system.

From a temporal design perspective, intensive longitudinal data (ILD) are necessary to address these key themes, but they also carry at least 5 challenges that influence the choice of the statistical model.

1) Differentiating between- and within-person effects.
2) ILD include time-varying variables (e.g. workloads) and may also incorporate stable (time-invariant) variables (e.g., sex).
3) The 'dependency' created by repeated measurements of the same individuals violates the assumption of 'independence' common to many traditional analyses [28,29].
4) Almost all longitudinal datasets have missing or unbalanced data [14].
5) Longitudinal data analysis require researchers to consider the role of time in their analysis [3].

Evaluating statistical approaches

We deliberately tried to align components 1 and 2 of Collins' framework by describing the theoretical models underpinning the workload—injury association and only including articles that had a temporal design characterised by ILD. To review whether statistical approaches aligned with these two components, two authors (JW + BZ) qualitatively assessed whether the statistical models, as employed in the included studies, (1) were multifactorial, (2) differentiated between- and within-athlete differences in injury risk, and (3) analyzed the data as a dynamic system – the three themes highlighted in the theoretical framework. From the temporal design, the same two authors evaluated whether the statistical analyses (4) included both time-varying and time-invariant variables, (5) were robust to missing/unbalanced data, 6) addressed the dependencies created by repeated measures, and (7) incorporated time into the analysis.

Data synthesis approach

We first describe the characteristics of the included articles, then present our qualitative assessment of how well the various statistical approaches fit within Collins' framework.

RESULTS

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. No item had more than 2 discrepancies. Of the 250 study criteria in the second set of 5 articles coded by both reviewers, there were 8 discrepancies (97% agreement).

Included articles were published from 2003 – 2016, with 78% of the studies published since 2010. Sports studied included rugby league (n = 10), soccer (n = 7), Australian football (n = 6), cricket (n = 5), rugby union (n = 2), multiple sports (n = 1), and basketball, handball, and volleyball (n = 1 each). Studies included an average of 96 athletes (median = 46), ranging from 12 [30] to 502 athletes [31]. The observation period for these cohort studies ranged from 14 weeks [32] to 6 years [33]. Most studies investigated male athletes (n = 30), with 2 studies on female athletes, and 2 on both sexes. Table 1 summarizes the included articles' basic characteristics, while the full data extraction table is available from the authors upon request.

Data collection

Injury definitions

Injury definitions varied across articles, with exact wording outlined in the Online Supplementary Appendix. In Table 2, we have categorised the definitions into more discrete injury categories (and subcategories) in accordance with recognized consensus statements [34].Where studies used multiple injury definitions, we categorized them according to the definition used for the primary analysis.

Reference	Journal	Sport	Study Length	Level	Sex	Age

Table 1: Summary of included workload—injury investigations, sorted by sport then publication year

(Rogalski et al., 2013) [35]	*J Sci Med Sport*	1 season	AFL	46	Elite	Male
(Colby et al., 2014) [36]	*JSCR*	1 Season	AFL	46	Professional	Male
(Duhig et al., 2016) [37]	*BJSM*	2 Seasons	AFL	51	Professional	Male
(Murray et al., 2016a) [38]	*Scand J Med Sci Sports*	2 Seasons	AFL	59	Professional	Male
(Murray et al., 2016b) [39]	*IJSPP*	1 season	AFL	46	Professional	Male
(Veugelers et al., 2015) [40]	*J Sci Med Sport*	15 weeks	AFL	45	Elite	Male
(Anderson et al., 2003) [30]	*JSCR*	21 weeks	Basketball	12	Sub-elite competitive	Female
(Dennis et al., 2003) [41]	*J Sci Med Sport*	2 Seasons	Cricket	90	Professional	Male
(Dennis et al., 2004) [42]	*J Sci Med Sport*	1 Season	Cricket	12	Professional	Male
(Dennis et al., 2005) [43]	*BJSM*	2002-2003 cricket season	Cricket	44	Sub-elite competitive	Male
(Saw et al., 2011) [44]	*BJSM*	1 season	Cricket	28	Elite	Male
(Hulin et al., 2014) [33]	*BJSM*	43 indiv. seasons / 6 years	Cricket	28	Professional	Male
(Bresciani et al., 2010) [45]	*Eur J Sport Sci*	1 season (40 weeks)	Handball	14	Elite	Male
(Gabbett, 2004a) [46]	*BJSM*	3 years	Rugby league	220	Sub-elite	Male
(Gabbett, 2004b) [47]	*J Sports Sci*	1 season	Rugby league	79	Semi-professional	Male
(Gabbett & Domrow, 2007) [48]	*J Sports Sci*	2 seasons	Rugby league	183	"Sub-elite"	Male
(Gabbett, 2010) [49]	*JSCR*	4 years	Rugby league	91	Professional	Male
(Gabbett & Jenkins, 2011) [50]	*J Sci Med Sport*	4 years	Rugby league	79	Professional	Male
(Gabbett & Ullah, 2012) [51]	*JSCR*	1 Season	Rugby league	34	Professional	Male
(Hulin et al., 2015) [52]	*BJSM*	2 Seasons	Rugby league	28	Professional	Male
(Hulin, et al., 2016) [53]	*BJSM*	2 Seasons	Rugby league	53	Professional	Male
(Windt et al., 2016) [54]	*BJSM*	1 season	Rugby league	30	Elite	Male
(Killen et al., 2010) [32]	*JSCR*	14 weeks	Rugby league	36	Professional	Male
(Brooks et al., 2008) [31]	*J Sports Sci*	2 seasons	Rugby Union	502	Professional	Male
(Cross et al., 2015) [55]	*IJSPP*	1 Season	Rugby Union	173	Professional	Male
(Arnason et al., 2004) [56]	*AJSM*	1 season	Soccer	306	Professional	Male
(Brink et al., 2010) [57]	*BJSM*	2 Seasons	Soccer	53	Elite Youth Players	Male
(Mallo & Dellal, 2012) [58]	*J Sports Med Phys Fitness*	2 seasons (2007/08 and 2008/09)	Soccer	35	Professional (Spanish Division II)	Male
(Clausen et al., 2014) [59]	*AJSM*	1 Season	Soccer	498	Recreational	Female
(Owen et al., 2015) [60]	*JSCR*	2 consecutive seasons	Soccer	23	Professional	Male
(Bowen et al., 2016) [61]	*BJSM*	2 Seasons	Soccer	32	Elite Youth Players	Male
(Ehrmann et al., 2016) [62]	*JSCR*	1 Season	Soccer	19	Professional	Male
(Malisoux et al., 2013) [63]	*J Sci Med Sport*	41 weeks	Varied	154	High-school	Both (65% males)
(Visnes & Bahr, 2013) [64]	*Scand J Med Sci Sports*	4 years (231 Student-seasons)	Volleyball	141	Elite high school	Both (72F, 69M)

Table 2: Broad injury definitions used in workload—injury investigations

Injury Definition	**N**
Time-Loss	
All time-loss	13
Match time-loss	2
Non-contact time-loss	7
Non-contact match time-loss	1
Medical Attention	
Medical attention	7
Player-reported pain, soreness, or discomfort	1
Non-contact medical attention injuries	1
Clinical diagnosis of jumper's knee	1
Other	
Injury scale on the Recovery-Stress Questionnaire for Athletes (REST-Q)	1

Subsequent or recurrent injuries

Of the 34 articles, 30 did not define or include subsequent or recurrent injuries. Of those that explicitly addressed subsequent injuries, two defined these injuries as those occurring at the same time and occurring by the same mechanism [58,60]. Two articles explicitly stated that they only considered time until first injury, meaning no injuries were subsequent or recurrent [41,43].

Workload definitions

Workload variables varied widely across articles and are summarized in Table 3. For a more detailed description of each article's load measures, see the Online Supplementary Appendix. Many articles used workload metrics to derive additional variables from workload distribution over time (e.g. monotony, strain, acute: chronic workload ratios).

Table 3: Independent variables used in workload—injury investigations

Workload Measure	**N**
Internal	
- sRPE	15
- Heart rate zones	2
External	
- Balls bowled or pitched	5
- GPS / accelerometry	10
- Hours	6

* If articles included more than one type of workload variable they are counted more than once. sRPE scores could be the original Foster scale or modified. GPS – Global positioning system. sRPE – Session-rating of perceived exertion (calculated as the product of session intensity on a 1-10 Borg Scale and activity duration in minutes).

Measurement frequency

Most included articles (n=32) collected workload data at every session that athletes completed, while 2 studies recorded workload on a weekly basis [59,64].

Handling missing data

Twenty-three of the 34 articles (67%) did not report any strategies for missing data. Of those that did, 5 used listwise or casewise deletion, and 6 used estimation. Estimation methods for players missing data included techniques such as: using the full team average values for the drills a player completed [61], using an individual's mean weekly value [57], and multiplying player's pre-season per-minute match data by the number of minutes they played in a match [36].

Statistical analysis and reporting in included articles

Data binning/aggregation

Although 32 articles collected daily workload measurements, many aggregated data for analysis. Most (n = 16) summed workload metrics for a total or average weekly workload. Three studies aggregated workload data for the entire year, 3 aggregated data into season periods, 2 aggregated data monthly, and 3 used multiple aggregation strategies.

Analysis methods

Table 4 summarises the statistical practices of applied researchers investigating the relationship between workload and injury. Although some studies had analysed other primary or secondary objectives, we recorded only the analyses used to investigate the workload – injury relationship.

Table 4: The number of studies using various statistical analysis techniques

Analytical Method	**N**
Regression modelling	
- Logistic	
○ Regular	10
○ Generalised Estimating Equation	5
○ Multilevel	1
- Linear	
○ Regular	2
- Poisson	
○ Generalised Estimating Equation**	1
- Multinomial regression	
○ Regular	1
- Cox proportional hazards model	1
- Frailty model	1
Correlation	
- Pearson	9
- Spearman	1
Relative risk/rate ratio*	8
T-tests	
- Paired and independent samples	4
- Independent samples only	2
Chi-square tests	1
Repeated measures ANOVA (one or two-way)	5

If articles used more than one statistical method to analyse workload and injury, they are included more than once in the table. We only report analyses used to analyse workload—injury associations, not other analyses reported in the articles (e.g. ANOVA to test for differences in total workloads at separate times of the season).
* Relative risk here refers to the use of RR as a primary analysis based on risks in different categorical groups, not as an effect estimated from another model. For example, comparing risks among different load groups like Hulin et al., (2014, 2016, 2016) are counted here, whereas Gabbett and Ullah (2012) derived RR from their frailty model, and Clausen et al. (2014) derived RR from their Poisson model, but neither are included in the count for

RR.
** Clausen et al. (2014) also report fitting multilevel models, but do not report any of the results – presenting only their GEE findings in their results and discussion.

Typical uses of statistical tools

Regression approaches were used most commonly (22/34 studies),. The most common approach was logistic regression (binary injury status as the outcome variable), independently or jointly modelling workload variables as independent variables. Generalised estimating equations were used in 5 studies to account for the clustering of observations within players and were used very similarly to simple logistic regression approaches.

Correlation was the second most common method (10/34 studies). Most studies that used correlation (7/10) measured the association between weekly or monthly workloads and injury incidence at the team level. Of those that used correlation at the individual level, two compared the number of completed pre-season sessions with the number of completed in-season sessions [23,65], while the final compared workload with injury operationalised as a numerical score on the injury subscale of the Recovery-Stress Questionnaire for Athletes (REST-Q) [45].

Relative risk approaches were generally used in one of two ways. First, workload categories were established for the entire year, like cricket bowlers who averaged <2, 2-2.99, 3-3.99, 4-4.99 or >5 days between bowling sessions up until an injury, or for the entire year if they did not sustain an injury [41]. Risks were calculated as the number of injuries/number of athletes in a given group, and relative risks were calculated to compare across groups [44]. In the second approach, athletes contributed exposures on a weekly basis, and thus contributed to multiple workload classifications. In this case, the likelihood/risk was the number of injuries/number of weekly player exposures to that workload category [33,53,61].

Group differences were sometimes evaluated using t-tests, ANOVAs or chi-square analyses. Typically, unpaired t-tests contrasted workload variables (e.g. mean sessions/week) between athletes who sustained an injury during the year, to those who did not [41,43]. Paired t-tests and repeated measures ANOVAs (one- or two-way) were most often used to contrast the workloads of the same athletes at different time periods. For example, workloads in an 'injury block' (like the week preceding an injury), were contrasted with non-injury blocks, like other weeks in the season [41,44], or the 4 weeks preceding the injury block [63].

Justifications for statistical approaches

Authors of 15 of the included articles (44%) did not cite any sources to support their analytical choices. Of those who did, most (n=14) cited previous literature in the sports medicine field. Eight articles referenced statistics or methodology articles, 4 cited articles on Prof. Will Hopkins' website (www.sportssci.org), and 3 cited statistical textbooks [66–68].

Addressing analysis assumptions and model fit

More than half (n = 20) the included articles did not report on the assumptions underlying their statistical analyses. Among those that did report on analysis assumptions, checks included checks for normality, collinearity of predictor variables in regression analyses [55], sphericity for repeated measures ANOVA [45], overdispersion [59], or correlation structures for generalised estimating equations [40].

When authors reported checking for normality, Shapiro-Wilk [65] or Kolmogorov-Smirnov tests [32] were referenced. Regression modelling was the most common analysis to investigate the workload—injury association. In 8/10 instances where simple logistic regressionwas chosen, the authors appear to have conducted the analyses using weekly observations without accounting for the dependencies created by repeated-measures across players. In all instances were regression was used, it was uncommon for authors to report that model assumptions were checked. Where multiple regression approaches were used, multicollinearity checks were rarely reported – an important consideration since multicollinearity can cause imprecise estimates of regression coefficients when multiple workload variables are simultaneously modelled [55,69,70].

Of the papers that modelled data using regression or similar techniques, six described how they assessed model fit. Some authors assessed specificity/sensitivity, or receiver operating characteristics, either on the current data set [64], or future data set [49]. Other in-sample model fit indices R^2 values [60], Aikeke Information Criteria

(AIC) and Bayesian Information Criteria (BIC), which were sometimes mentioned as guiding the model selection process [54].

Alignment of authors' statistical models with theoretical model and temporal design challenges

In Table 5 (a more detailed table – Table A2 - is available in Appendix 2), we qualitatively evaluated whether the statistical approaches chosen by the authors in our current review effectively addressed the key themes/challenges presented by the theoretical model and the temporal design (intensive longitudinal data). This table is an analytical tool to guide the reader through the discussion. It highlights the themes/challenges of the theoretical model and temporal design, as well as the strengths/weaknesses of the statistical tools used in included studies. The table has the challenges/themes in columns and statistical tools in rows. The reader can follow a row to see how well a given statistical tool addressed key challenges as used by researchers in our included articles, or they can choose a challenge and follow the column down to see which analyses were used in a way that addressed that challenge adequately. The rows are ordered according to their qualitative 'score'. As one proceeds down the rows, the statistical tools address more of the temporal design and theoretical model challenges.

We caution the reader that (1) not every possible statistical tool is included in the table, only those used in at least 1 article in our review, and (2) the evaluation is based on whether researchers of our included papers used a test in a way that addressed a given challenge, not necessarily whether the test is capable of being used in a way that meets that challenge. For example, a logistic regression analysis conducted using a generalised estimating equation framework can include multiple explanatory/predictor variables, thereby allowing for a multifactorial model. However, some authors used GEEs and only included one predictor variable [42,48], in which case the GEE did not address the multifactorial theme.

DISCUSSION

We used the workload—injury field of medical research to examine whether statistical approaches analyse intensive longitudinal data optimally. By design, the theoretical models underpinning the workload—injury field and the temporal design (ILD) were aligned in all the included articles, but common statistical approaches varied in how adequately they addressed the key themes needed to align them with the other two components.

Consideration #1 – Theoretical theme – multifactorial aetiology

Sports injury aetiology models of the last 2 decades have highlighted the multifactorial nature of athletic injury. [19,21]. We asked whether the burgeoning body of research relating workloads and injury, is using modern statistical methods to capture workloads while incorporating known risk factors. Few articles in this review incorporated previously identified risk factors and workload into the same analysis. In some instances, the analytical approach prevented this from being an option. For example, simple analyses like t-tests, correlations, and chi-square tests do not allow for multiple variables to be included. In other instances, the statistical approaches allowed a multifactorial approach (e.g. generalised estimating equations) but researchers opted to focus on the effects of workloads in isolation [42,48].

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 [71,72]. 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 [73,74]. As a reminder, there are subtle, but important differences between mediation, moderation, and effect measure modification that will influence analytical choices [75,76]. 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.

Table 5: Evaluation of the degree to which authors' use of statistical tools addressed theoretical and temporal design challenges

		Themes of theoretical model			Themes of temporal design - intensive longitudinal data			
Method	n	Multifactorial aetiology	Between and Within-Athlete Differences	Complex System	Includes Time-Varying and Time-Invariant Variables	Missing/ Unbalanced Data*	Repeated Measure Dependency	Incorporates Time into the Analysis
Correlation (Pearson and Spearman)	10	X	X	X	X	X	X	X
Unpaired t-test	6	X	X	X	X	X	X	X
Chi-square tests	1	X	X	X	X	X	X	X
Relative risk calculations	8	O	X	X	X	X	X	X
Regression (logistic, linear, multinomial)	13	O	X	X	X	X	X	X
Paired t-test	2	X	X	X	X	X	✓	✓
Repeated measures ANOVA (one or two-way)	5	O	O	X	O	X	✓	✓
Generalised Estimating Equations (Poisson and logistic)	6	O	X	X	O	✓	✓	O
Cox proportional hazards model	1	✓	X	X	X	✓	✓	✓
Multilevel modeling	1	✓	✓	X	✓	✓	✓	X
Frailty model	1	✓	✓	X	✓	✓	✓	✓

Qualitative assessment performed on a three-tiered scale. An 'X' (red formatting) means that none of the authors using this tool adequately addressed that specific challenge. In some cases, this may be because the statistical model was unable to address it, and other times it may be because of the way they used it. An 'O' (yellow formatting) indicates that some authors addressed that challenge while others did not. This generally happened when the statistical tool could address a challenge but the authors sometimes chose not to use it in that way. A '✓' (green formatting) indicates that all authors using this statistical tool addressed that challenge adequately. *Missing/unbalanced data here is that caused by intensive longitudinal data – meaning a different number of observations for each athlete during the observation period, some of which may be missing.

Statistical approaches that allow multivariable analyses enabled researchers to examine the effects of workloads while controlling for known risk factors. Malisoux et al. (2013) used a Cox-proportional hazards model to control for age and sex while examining the effects of average training volume and intensity. Gabbett and Ullah's frailty model (2012) incorporated previous injury – a proven injury risk factor – into the evaluation of the influence of different GPS workloads on injury risk. When investigating multifactorial phenomena, statistical approaches that enable multiple explanatory variables provide a more appropriate option.

Consideration #2 – Theoretical theme - between and within-athlete differences

One of the primary benefits of ILD is that it enables researchers (when using certain analyses) to differentiate within-person and between-person effects [3]. In the sports medicine field, this would correspond to researchers asking, (1) why do some athletes suffer few injuries (between-person inquiry) while others appear 'injury-prone'? and on the other hand, (2) at what point is a given athlete (within-person inquiry) more likely to sustain an injury? The simpler statistical approaches used by researchers in our included studies (correlation, t-tests, ANOVAs, regular regression) are limited in the number of variables they can include, and consequently cannot differentiate risk between- and within-athletes. Tests of group differences (independent sample t-tests and one-way ANOVAs) only differentiate between athletes (e.g. injured vs. uninjured), while repeated measures tests (repeated measures ANOVA and paired t-tests) only examine within-athlete differences (e.g. loads preceding injuries vs. loads during non-injury weeks).

Generalised estimating equations (GEE) were commonly used to address some of the longitudinal data challenges. Although this approach accounts for the clustering within-persons, it assumes the effects of predictor variables are constant across all athletes [77]. Simple Cox proportional hazards models (Malisoux et al. 2013) are common in survival analyses, but do not differentiate between- and within-person effects [78].

Only two statistical tools were used in a way that examined between- and within-athlete differences in injury risk. The frailty model by Gabbett and Ullah (2012) modelled each athlete as a random effect with a given frailty. The multilevel model by Windt et al., (2016) incorporated athlete-level variables (age, position, pre-season sessions) and observation-level variables (weekly workload measures). In the latter case, athletes' weekly distances did not affect their risk of injury in the subsequent week (OR = 0.82 for 1 standard deviation increase, 95% CI = 0.55 to 1.21) – a within-athlete inquiry. However, controlling for weekly distance and the proportion of distance at high speeds, athletes who had completed a greater number of pre-season training sessions had significantly reduced odds of injury (OR = 0.83 for each 10 pre-season sessions, 95% CI = 0.70 to 0.99) – a between-athlete inquiry. These two examples highlight that certain analyses carry a distinct advantage of allowing researchers to tease out differences both between- and within- study participants

Consideration #3 – Theoretical theme - injury risk as a complex dynamic system

Complex systems are defined, among other things, by the interaction between multiple internal and external variables that interact to produce an outcome. Simple analyses (t-tests, correlations), which cannot incorporate multiple variables, cannot examine the interaction between multiple factors. However, even other traditional analyses which are more effective in handling the challenges of longitudinal data (e.g. generalised estimating equations, Cox-proportional hazards models) were not used to incorporate non-linear interactions between predictor variables.

The most recent reviews of athletic injury aetiology have highlighted complex systems models [24,79]. None of the analyses included in our review analysed intensive longitudinal workload—injury data with statistical analyses that fit within a complex systems framework. This lack of research may reflect the fact that the suggestion that injury aetiology fits within a dynamic, complex systems framework is still relatively ‘new’ in this field. It remains to be seen whether a complex systems approach and the analyses recommended in such reviews (e.g. self-organising feature maps, classification and regression trees, agent based models, etc.) are more effective for evaluating the association between workloads and injury [24].

Consideration #4 – ILD challenge – including time-varying and time-invariant variables

Tying back to the theoretical model of workloads and injury, some relevant factors may be relatively stable (time-invariant) over the course of an observation period (e.g. height, age), while others are time-varying (e.g. workload). Some analyses can incorporate both time-varying and time-invariant variables, while others are limited in this respect. All analyses that cannot or did not address the multifactorial nature of injury cannot include time-varying and time-invariant variables concurrently. Group difference tests (t-tests, ANOVAs, etc.) may collect time-varying measures, but must aggregate them into a single average for analysis.

Including time-varying and stable variables in the same analysis links closely to between- and within-athlete differences – with the frailty model [51] and multilevel model [54] both used in a way that allowed the researchers to include both. The one exception in our included studies was the generalised estimating equation approach. As mentioned earlier, the GEE assumes an 'overall' effect for each explanatory variable, such that between- and within-athlete differences cannot be differentiated [80]. However, the major benefit to a GEE is that it accounts for the repeated measures for each participant and can therefore include both time-invariant and time-varying variables for each participant.

Consideration #5 – ILD challenge – handling missing and unbalanced data

Dealing with missing and unbalanced data is a near certainty when collecting ILD, and is common in applied workload-monitoring settings [81]. 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 [82,83]. 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 [84].

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 [85]. 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 [85].

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.) [33,61], '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 [82,84].

Challenge #6 – ILD challenge – dependencies created by repeated measures

Collecting ILD in applied sport settings means repeated (often daily) measurements of the same athlete – such that observations are clustered within athletes. Comparisons of independent groups, through chi-square tests, independent sample t-tests and one-way/two-way ANOVAs all assume participants contribute a single observation to the analysis and force an aggregated variable (e.g. average number of balls bowled in a week) to conduct the analysis [41]. Similarly, correlation and simple regression (in its linear, logistic, and multinomial forms) assume independence of observations [86]. Paired t-tests and repeated measures ANOVA were used to deal with repeated measurements by comparing the same athletes' workloads at different periods (e.g. the week before injury vs. weeks that did no precede injury).

Of the analyses that addressed this challenge, GEEs were used most commonly (6 studies). GEE's ability to handle clustering was also used in one article to control for players clustering within teams [59]. Cox proportional hazard models, used in one article [63], can handle repeated measurements for participants [87]. Multilevel models [54] and frailty models [51] – an extension of the Cox- proportional hazards model – were also used in a single instance each, where repeated measures were clustered within players through a random player effect.

As mentioned in our introduction, there may be additional data dependency created by recurrent injuries [88]. Previous recommendations to handle the recurrent injury challenges have included frailty models [89], and a multistate framework [90]. However, as so few articles reported collecting information on recurrent injuries (n = 5), we focused primarily on the dependencies caused by repeated measures across participants.

Challenge #7 – ILD challenge – incorporating time into the analysis / temporality

One of the most relevant questions in ILD analyses is the way that time is accounted for [1,3]. Some authors used one-way [32] and two-way repeated measures ANOVAs [46] to compare loading in different seasons or season-periods – a very simple way of accounting for time. Repeated measures ANOVAs [62,63,65] and paired t-tests [41,44] also account for time by categorising time-periods as pre-injury blocks or non-injury blocks. Multilevel models have been used to examine change through the interactions of variables with time, but the one multilevel model used in this review did not include time as a covariate [54]. Survival analyses explicitly account for time by calculating the effects of variables on the predicted time-to-event [51,63]. Notably, only one analysis – the frailty model (Gabbett and Ullah, 2012) – adjusted the probability of long-term outcomes (e.g., injury) based on variations after an initial capture of risk, something few traditional analyses accomplish (Cook 2016).

Temporality is also vital in considering potential causal associations. While making causal inferences from observational data is a topic beyond the scope of this paper, temporality is a well-accepted component of causality dating back, at least, to Bradford Hill's 'criteria' [91]. Without temporality – where a postulated cause precedes the outcome – directional associations cannot be made [92,93]. A lack of temporality can also skew associations since it allows for reverse causality. In the workload—injury field, findings that high weekly workloads are sometimes associated with lower odds of injury in a given week [40,54] may be in part because players who get injured in a given week are less likely to accumulate high weekly workloads.

Trying to account for temporality, some researchers have included a latent period – where workload variables are examined for their association with injury occurrence in a given proceeding time window, like the subsequent week [33,54]. While recent work has noted that the length of the latent period may affect model findings [94], it is clear that without some type of latent period, any directional inferences between workloads and injury cannot be made.

Methodological, statistical and reporting considerations

Data aggregation

Data aggregation was common, whether in data preparation, or forced through the analysis. In some cases, researchers aggregated individual level data into team-level measures (total/average workload and injury incidence). Although 32/34 articles collected daily data, most aggregated these daily data into weekly measures, potentially contributing to temporality problems if no latent period was included. Finally, certain analyses (e.g. paired t-tests, simple logistic regression) aggregated data for athletes across an entire year so that workload measures were used to control for exposure [56,64]. Differences in analyses make it impossible to measure the effect of fluctuations in workload and potential impact on injury risk. Further, with no latent period, the directionality of the relationship is unclear. For example, players with high exposure throughout the year were at a lower injury risk than the intermediate group, but it could be interpreted that players who do not sustain an injury throughout the year are more likely to accumulate high total training and match exposures (i.e. higher workload) [56]. Aggregated data may be easier to analyse but comes at the cost of losing some of the inherent benefits of collecting ILD, such as the changes in injury risk that occur at a daily level. As a result, theory-driven questions that relate to daily workload fluctuations and injury risk will become challenging, or impossible to answer.

Checking model assumptions and fit

While many studies may have under-reported how they assessed model assumptions or fit, others [55] provide an example for other researchers to emulate. In fitting a GEE to account for intra-team and intra-player clustering effects, they explained how they selected an appropriate autocorrelation structure, reported how potential quadratic relationships were assessed in the case of non-linear associations, and described checking for potential multicollinearity with defined thresholds (variance inflation factor >10) and for their GEE.

Researcher 'trade-offs', consequences of misalignment

We used the workload—injury field to highlight seven themes that relate to theoretical injury aetiology models and temporal design (ILD). In many caseshighlighted published studies' statistical models either could not, or were not used in a way that addresses these themes. In some cases, misalignment may carry a severe cost – like assumption violations that may bias study results [29]. This is akin to building conclusions on an unstable foundation. Other times, researchers have properly employed their chosen statistical approach, but the approaches themselves were limited, and unable to answer research questions that ILD can address. This is more akin to having a grand building plan and all the necessary supplies, but only using a screwdriver to construct the building.

Simple regression models provide an ideal example of researchers' trade-offs when using traditional statistical analyses on ILD, and the potential costs of misalignment. Although 13 papers used regular regression to analyse the association between workloads and injury outcome, they chose one of three paths when dealing with ILD. First, many proceeded to analyse each daily or weekly data point as an independent observation – not addressing the violation of the independence assumption [33,53,95]. Second, some researchers aggregated the workload data into an average weekly workload or total workload exposure over the course of the year, such that each participant contributed only one observation to a classic logistic regression [56,64]. Although the regression assumptions were not violated, workload was aggregated into a single metric, the temporal relationship between workload and injury was lost, and there was then no way to analyse the effects of workload fluctuations on injury risk. Third, some researchers converted individual data to team level data and examined team workloads with team injury incidence in a linear regression [47,60]. In this final case, no differentiation could then be made between players or within-players, and inferences were only possible at the team level. This may be sufficient to inform research on the association of workloads and injury at the team level, but the theoretical model underpinning team injury rates may differ from those that underpin individual athletes' injury risk.

Review limitations

Previous systematic reviews investigating the workload—injury relationship have documented the challenges of identifying articles through classic systematic review search strategies [7,9]. Heterogeneous keywords and the breadth of sporting contexts have meant previous systematic reviews include many articles post-hoc that were not originally identified by their systematic searches (e.g. 29 of 67 articles in Jones [7], 12 of 35 articles in the paper by Drew and Finch [9]). Therefore, although we worked to identify articles through 6 systematic reviews [7–10,17,18] and the 2016 IOC consensus statement on athletic workloads and injury [11], we may have missed potentially eligible articles .

We used the cut-off for intensive longitudinal data (>20 observations) proposed by Collins (2006). However, there is no universal cut-off for ILD, with previous thresholds of 'more than a handful' [1], ten observations [96], or forty [3].

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.

Finally, it was beyond the scope of this review to list every challenge posed by ILD, and we were not exhaustive in our discussion of different analyses and their capacity to handle the challenges. Where possible, we tried to identify the themes that are most common within the research field of Sport and Exercise Medicine field. Ultimately, our call to action is that statistical tools be chosen more thoughtfully so that the extensive work put into theory building and data collection is not short-changed by a sub-optimal statistical model.

Longitudinal improvements in ILD analysis

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 [97]. 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 [98–103].

Mediation, effect measure modification, and interaction/moderation are all causal models which may also contribute to aetiological frameworks [104]. We recently proposed that traditional intrinsic and extrinsic risk factors may act as moderators or effect measure modifiers of the workload—injury association [73]. 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 [74]. 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 [100,105,106]. 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) [76]. 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).

Future directions and recommendations for ILD analysis

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 [107]. Because of sample structure, mixed models prevent false positive associations and have an applied correction method that increases the power of the analysis [108]; 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) [109]. 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 [109]. For more information on multilevel/mixed effect models for longitudinal analysis, readers are referred to a other helpful resources [1,28,85,110,111].

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 [112] – 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 [113–115]. 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 [113]. 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 [116]. 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 [87,117,118].

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[119], and open the door to more complex systems modelling (e.g. agent-based modelling) [97]. Though they show promise, such simulation studies are based on artificially generated data and must be interpreted carefully [120].

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 [121], machine learning approaches [98,101], time series analysis [122], and time-varying effect models [123] 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 [110].

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 [124]. 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.

CONCLUSION

We used studies investigating the relationship between workloads and injury as a substrate to highlight to researchers how important it is to align their theoretical model, temporal design, and statistical model. In longitudinal research, thoughtfully chosen statistical analyses are those grounded in subject matter theory and that maximise the utility of the collected data. The three most common analyses in our included papers (logistic regression, correlations, and relative risk calculations) addressed one or none of the 3 key theoretical themes, and one or fewer of the 4 inherent challenges of intensive longitudinal data. In this example discipline, researchers have developed sophisticated theories and frequently collect data that enable them to test these theoretical models. The missing step, and future opportunity for researchers, is to avail themselves of all the tools at their disposal – choosing statistical models that address the ILD challenges and answer theory-driven research questions.

Contributions: JW and TG searched for, screened, and identified appropriate systematic reviews and consensus statements. JW identified appropriate original data papers from included systematic reviews/consensus statements. JW and CA performed quality assessment of original data papers and coded the study methods. BZ provided the statistical feedback on the creation of the data extraction spreadsheet and advice on the approach of the methodological review. JW and BZ completed the qualitative assessment of whether authors' use of statistical tools aligned with theoretical or temporal design themes. BS, CA, BZ and KK provided early input during early stages of study development CC contributed to the original idea for the review, and contributed to the discussion section of the manuscript. JW compiled the first draft of the manuscript. CA, CC, BS, BZ, and KK all contributed to critical revision of multiple drafts of the current manuscript.

Acknowledgements: The United States Coalition for the Prevention of Illness & Injury in Sport is a partner in the United States Coalition for the Prevention of Illness and Injury in Sport an International Research Centres for Prevention of Injury and Protection of Athlete Health supported by the International Olympic Committee (IOC).

Technical or equipment support for this study was not provided by any outside companies, manufacturers or organizations.

Funding: JW was a Vanier Scholar funded by the Canadian Institutes of Health Research.

Competing interests: None

Provenance and peer review: Not commissioned, externally peer reviewed.

Ethics approval: No ethics approval was sought for conducting this methodological review.

Data sharing statement: We do not have any original data in the current methodological review to share.

REFERENCES

1 Walls TA, Schafer JL. *Models for intensive longitudinal data*. Oxford [U.K.]: : Oxford University Press 2006. http://www.loc.gov/catdir/enhancements/fy0636/2005047269-t.html (accessed 13 Apr 2016).

2 Ginexi EM, Riley W, Atienza AA, *et al.* The promise of intensive longitudinal data capture for behavioral health research. *Nicotine Tob Res Off J Soc Res Nicotine Tob* 2014;**16 Suppl 2**:S73–5. doi:10.1093/ntr/ntt273

3 Hamaker EL, Wichers M. No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Curr Dir Psychol Sci* 2017;**26**:10–5. doi:10.1177/0963721416666518

4 Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. *Sports Med* 2014;**44**:139–47. doi:10.1007/s40279-014-0253-z

5 McCall A, Carling C, Nedelec M, *et al.* Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med* 2014;**48**:1352–7. doi:10.1136/bjsports-2014-093439

6 Akenhead R, Nassis GP. Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *Int J Sports Physiol Perform* Published Online First: 9 October 2015. doi:10.1123/ijspp.2015-0331

7 Jones CM, Griffiths PC, Mellalieu SD. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Med* 2016;:1–32. doi:10.1007/s40279-016-0619-5

8 Black GM, Gabbett TJ, Cole MH, *et al.* Monitoring Workload in Throwing-Dominant Sports: A Systematic Review. *Sports Med* 2016;:1–14. doi:10.1007/s40279-016-0529-6

9 Drew MK, Finch CF. The Relationship Between Training Load and Injury, Illness and Soreness: A Systematic and Literature Review. *Sports Med* 2016;:1–23. doi:10.1007/s40279-015-0459-8

10 Gabbett TJ, Whyte DG, Hartwig TB, *et al.* The Relationship Between Workloads, Physical Performance, Injury and Illness in Adolescent Male Football Players. *Sports Med* 2014;**44**:989–1003. doi:10.1007/s40279-014-0179-5

11 Soligard T, Schwellnus M, Alonso J-M, *et al.* How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med* 2016;**50**:1030–41. doi:10.1136/bjsports-2016-096581

12 Keselman HJ, Huberty CJ, Lix LM, *et al.* Statistical Practices of Educational Researchers: An Analysis of Their ANOVA, MANOVA, and ANCOVA Analyses. *Rev Educ Res* 1998;**68**:350–86. doi:10.2307/1170601

13 Weir A, Rabia S, Ardern C. Trusting systematic reviews and meta-analyses: all that glitters is not gold! *Br J Sports Med* 2016;:bjsports-2015-095896. doi:10.1136/bjsports-2015-095896

14 Collins L. Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model. *Annu Rev Psychol* 2006;**57**:505–28. doi:10.1146/annurev.psych.57.102904.190146

15 Lemon SC, Wang ML, Haughton CF, *et al.* Methodological quality of behavioural weight loss studies: a systematic review. *Obes Rev* 2016;:n/a-n/a. doi:10.1111/obr.12412

16 Gabbett TJ, Nassis GP, Oetter E, *et al.* The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data. *Br J Sports Med* 2017;:bjsports-2016-097298. doi:10.1136/bjsports-2016-097298

17 Olivier B, Taljaard T, Burger E, *et al.* Which Extrinsic and Intrinsic Factors are Associated with Non-Contact Injuries in Adult Cricket Fast Bowlers? *Sports Med* 2015;:1–23. doi:10.1007/s40279-015-0383-y

18 Whittaker JL, Small C, Maffey L, *et al.* Risk factors for groin injury in sport: an updated systematic review. *Br J Sports Med* 2015;**49**:803–9. doi:10.1136/bjsports-2014-094287

19 Bahr R, Holme I. Risk factors for sports injuries — a methodological approach. *Br J Sports Med* 2003;**37**:384–92. doi:10.1136/bjsm.37.5.384

20 Bahr R, Krosshaug T. Understanding injury mechanisms: a key component of preventing injuries in sport. *Br J Sports Med* 2005;**39**:324–9. doi:10.1136/bjsm.2005.018341

21 Meeuwisse WH. Assessing Causation in Sport Injury: A Multifactorial Model. *Clin J Sport Med* 1994;**4**:166–170.

22 Meeuwisse WH, Tyreman H, Hagel B, *et al.* A Dynamic Model of Etiology in Sport Injury: The Recursive Nature of Risk and Causation. *Clin J Sport Med* 2007;**17**:215–9. doi:10.1097/JSM.0b013e3180592a48

23 Windt J, Gabbett TJ. How do training and competition workloads relate to injury? The workload—injury aetiology model. *Br J Sports Med* 2016;:bjsports-2016-096040. doi:10.1136/bjsports-2016-096040

24 Bittencourt NFN, Meeuwisse WH, Mendonça LD, *et al.* Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *Br J Sports Med* 2016;:bjsports-2015-095850. doi:10.1136/bjsports-2015-095850

25 Hulme A, Salmon PM, Nielsen RO, *et al.* Closing Pandora's Box: adapting a systems ergonomics methodology for better understanding the ecological complexity underpinning the development and prevention of running-related injury. *Theor Issues Ergon Sci* 2017;**18**:338–59. doi:10.1080/1463922X.2016.1274455

26 Cook C. Predicting future physical injury in sports: it's a complicated dynamic system. *Br J Sports Med* 2016;:bjsports-2016-096445. doi:10.1136/bjsports-2016-096445

27 Sun L, Wu R. Mapping complex traits as a dynamic system. *Phys Life Rev* 2015;**13**:155–85. doi:10.1016/j.plrev.2015.02.007

28 Hoffman L. *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. Routledge 2014. http://www.tandfebooks.com/isbn/9781315744094 (accessed 8 Apr 2016).

29 Wilkinson M, Akenhead R. Violation of statistical assumptions in a recent publication? *Int J Sports Med* 2013;**34**:281. doi:10.1055/s-0032-1331775

30 Anderson L, Triplett-McBride T, Foster C, *et al.* Impact of training patterns on incidence of illness and injury during a women's collegiate basketball season. *J Strength Cond Res* 2003;**17**:734–8.

31 Brooks JHM, Fuller CW, Kemp SPT, *et al.* An assessment of training volume in professional rugby union and its impact on the incidence, severity, and nature of match and training injuries. *J Sports Sci* 2008;**26**:863–73. doi:10.1080/02640410701832209

32 Killen NM, Gabbett TJ, Jenkins DG. Training Loads and Incidence of Injury During the Preseason in Professional Rugby League Players: *J Strength Cond Res* 2010;**24**:2079–84. doi:10.1519/JSC.0b013e3181ddafff

33 Hulin BT, Gabbett TJ, Blanch P, *et al.* Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *Br J Sports Med* 2014;**48**:708–12. doi:10.1136/bjsports-2013-092524

34 Fuller CW, Ekstrand J, Junge A, *et al.* Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med* 2006;**40**:193–201. doi:10.1136/bjsm.2005.025270

35 Rogalski B, Dawson B, Heasman J, *et al.* Training and game loads and injury risk in elite Australian footballers. *J Sci Med Sport* 2013;**16**:499–503. doi:10.1016/j.jsams.2012.12.004

36 Colby MJ, Dawson B, Heasman J, *et al.* Accelerometer and GPS-Derived Running Loads and Injury Risk in Elite Australian Footballers: *J Strength Cond Res* 2014;**28**:2244–52. doi:10.1519/JSC.0000000000000362

37 Duhig S, Shield AJ, Opar D, *et al.* Effect of high-speed running on hamstring strain injury risk. *Br J Sports Med* 2016;:bjsports-2015-095679. doi:10.1136/bjsports-2015-095679

38 Murray NB, Gabbett TJ, Townshend AD, *et al.* Individual and combined effects of acute and chronic running loads on injury risk in elite Australian footballers. *Scand J Med Sci Sports* 2016;:n/a-n/a. doi:10.1111/sms.12719

39 Murray NB, Gabbett TJ, Townshend AD. Relationship Between Pre-Season Training Load and In-Season Availability in Elite Australian Football Players. *Int J Sports Physiol Perform* 2016;:1–21. doi:10.1123/ijspp.2015-0806

40 Veugelers KR, Young WB, Fahrner B, *et al.* Different methods of training load quantification and their relationship to injury and illness in elite Australian football. *J Sci Med Sport* 2015;**19**:24–8. doi:10.1016/j.jsams.2015.01.001

41 Dennis R, Farhart R, Goumas C, *et al.* Bowling workload and the risk of injury in elite cricket fast bowlers. *J Sci Med Sport* 2003;**6**:359–67. doi:10.1016/S1440-2440(03)80031-2

42 Dennis R, Farhart P, Clements M, *et al.* The relationship between fast bowling workload and injury in first-class cricketers: a pilot study. *J Sci Med Sport Sports Med Aust* 2004;**7**:232–6.

43 Dennis RJ, Finch CF, Farhart PJ. Is bowling workload a risk factor for injury to Australian junior cricket fast bowlers? *Br J Sports Med* 2005;**39**:843–6. doi:10.1136/bjsm.2005.018515

44 Saw R, Dennis RJ, Bentley D, *et al.* Throwing workload and injury risk in elite cricketers. *Br J Sports Med* 2011;**45**:805–8. doi:10.1136/bjsm.2009.061309

45 Bresciani G, Cuevas MJ, Garatachea N, *et al.* Monitoring biological and psychological measures throughout an entire season in male handball players. *Eur J Sport Sci* 2010;**10**:377–84. doi:10.1080/17461391003699070

46 Gabbett TJ. Reductions in pre-season training loads reduce training injury rates in rugby league players. *Br J Sports Med* 2004;**38**:743–9. doi:10.1136/bjsm.2003.008391

47 Gabbett TJ. Influence of training and match intensity on injuries in rugby league. *J Sports Sci* 2004;**22**:409–17. doi:10.1080/02640410310001641638

48 Gabbett TJ, Domrow N. Relationships between training load, injury, and fitness in sub-elite collision sport athletes. *J Sports Sci* 2007;**25**:1507–19. doi:10.1080/02640410701215066

49 Gabbett TJ. The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *J Strength Cond Res* 2010;**24**:2593–603. doi:10.1519/JSC.0b013e3181f19da4

50 Gabbett TJ, Jenkins DG. Relationship between training load and injury in professional rugby league players. *J Sci Med Sport* 2011;**14**:204–9. doi:10.1016/j.jsams.2010.12.002

51 Gabbett TJ, Ullah S. Relationship Between Running Loads and Soft-Tissue Injury in Elite Team Sport Athletes. *J Strength Cond Res* 2012;**26**:953–60. doi:10.1519/JSC.0b013e3182302023

52 Hulin BT, Gabbett TJ, Lawson DW, *et al.* The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med* 2015;:bjsports-2015-094817. doi:10.1136/bjsports-2015-094817

53 Hulin BT, Gabbett TJ, Caputi P, *et al.* Low chronic workload and the acute:chronic workload ratio are more predictive of injury than between-match recovery time: a two-season prospective cohort study in elite rugby league players. *Br J Sports Med* 2016;:bjsports-2015-095364. doi:10.1136/bjsports-2015-095364

54 Windt J, Gabbett TJ, Ferris D, *et al.* Training load--injury paradox: is greater preseason participation associated with lower in-season injury risk in elite rugby league players? *Br J Sports Med* 2016;:bjsports-2016-095973. doi:10.1136/bjsports-2016-095973

55 Cross MJ, Williams S, Trewartha G, *et al.* The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *Int J Sports Physiol Perform* Published Online First: 26 August 2015. doi:10.1123/ijspp.2015-0187

56 Arnason A, Sigurdsson SB, Gudmundsson A, *et al.* Risk Factors for Injuries in Football. *Am J Sports Med* 2004;**32**:5S-16S. doi:10.1177/0363546503258912

57 Brink MS, Nederhof E, Visscher C, *et al.* Monitoring load, recovery, and performance in young elite soccer players. *J Strength Cond Res Natl Strength Cond Assoc* 2010;**24**:597–603. doi:10.1519/JSC.0b013e3181c4d38b

58 Mallo J, Dellal A. Injury risk in professional football players with special reference to the playing position and training periodization. *J Sports Med Phys Fitness* 2012;**52**:631–8.

59 Clausen MB, Zebis MK, Møller M, *et al.* High Injury Incidence in Adolescent Female Soccer. *Am J Sports Med* 2014;**42**:2487–94. doi:10.1177/0363546514541224

60 Owen AL, Forsyth JJ, Wong DP, *et al.* Heart rate-based training intensity and its impact on injury incidence among elite-level professional soccer players. *J Strength Cond Res Natl Strength Cond Assoc* 2015;**29**:1705–12. doi:10.1519/JSC.0000000000000810

61 Bowen L, Gross AS, Gimpel M, *et al.* Accumulated workloads and the acute:chronic workload ratio relate to injury risk in elite youth football players. *Br J Sports Med* 2016;:bjsports-2015-095820. doi:10.1136/bjsports-2015-095820

62 Ehrmann FE, Duncan CS, Sindhusake D, *et al.* GPS and Injury Prevention in Professional Soccer. *J Strength Cond Res* Published Online First: 2015. doi:10.1519/JSC.0000000000001093

63 Malisoux L, Frisch A, Urhausen A, *et al.* Monitoring of sport participation and injury risk in young athletes. *J Sci Med Sport* 2013;**16**:504–8. doi:10.1016/j.jsams.2013.01.008

64 Visnes H, Bahr R. Training volume and body composition as risk factors for developing jumper's knee among young elite volleyball players. *Scand J Med Sci Sports* 2013;**23**:607–13. doi:10.1111/j.1600-0838.2011.01430.x

65 Murray NB, Gabbett TJ, Townshend AD, *et al.* Individual and combined effects of acute and chronic running loads on injury risk in elite Australian footballers. *Scand J Med Sci Sports* 2017;**27**:990–8. doi:10.1111/sms.12719

66 Cohen J. *Statistical power analysis for the behavioral sciences. 2nd*. Hillsdale, NJ: erlbaum 1988.

67 Kraemer HC, Thiemann S. *How many subjects*. Citeseer 1987.

68 Kutner MH, Nachtsheim C, Neter J. *Applied linear regression models*. McGraw-Hill/Irwin 2004.

69 Silvey SD. Multicollinearity and Imprecise Estimation. *J R Stat Soc Ser B Methodol* 1969;**31**:539–52.

70 Zidek JV, Wong H, Le ND, *et al.* Causality, measurement error and multicollinearity in epidemiology. *Environmetrics* 1996;**7**:441–451.

71 Datson N, Hulton A, Andersson H, *et al.* Applied Physiology of Female Soccer: An Update. *Sports Med* 2014;**44**:1225–40. doi:10.1007/s40279-014-0199-1

72 Hewett TE, Myer GD, Ford KR. Anterior Cruciate Ligament Injuries in Female Athletes: Part 1, Mechanisms and Risk Factors. *Am J Sports Med* 2006;**34**:299–311. doi:10.1177/0363546505284183

73 Windt J, Zumbo BD, Sporer B, *et al.* Why do workload spikes cause injuries, and which athletes are at higher risk? Mediators and moderators in workload–injury investigations. *Br J Sports Med* 2017;:bjsports-2016-097255. doi:10.1136/bjsports-2016-097255

74 Nielsen RO, Bertelsen ML, Møller M, *et al.* Training load and structure-specific load: applications for sport injury causality and data analyses. *Br J Sports Med* 2017;:bjsports-2017-097838. doi:10.1136/bjsports-2017-097838

75 Corraini P, Olsen M, Pedersen L, *et al.* Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. *Clin Epidemiol* 2017;**9**:331–8. doi:10.2147/CLEP.S129728

76 VanderWeele TJ. On the Distinction Between Interaction and Effect Modification: *Epidemiology* 2009;**20**:863–71. doi:10.1097/EDE.0b013e3181ba333c

77 Ghisletta P, Spini D. An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. *J Educ Behav Stat* 2004;**29**:421–37. doi:10.3102/10769986029004421

78 Bellera CA, MacGrogan G, Debled M, *et al.* Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol* 2010;**10**:20. doi:10.1186/1471-2288-10-20

79 Hulme A, Finch CF. From monocausality to systems thinking: a complementary and alternative conceptual approach for better understanding the development and prevention of sports injury. *Inj Epidemiol* 2015;**2**. doi:10.1186/s40621-015-0064-1

80 Ward P, Coutts AJ, Pruna R, *et al.* Putting the 'i' Back in Team. *Int J Sports Physiol Perform* 2018;:1–14. doi:10.1123/ijspp.2018-0154

81 Buchheit M. Applying the acute:chronic workload ratio in elite football: worth the effort? *Br J Sports Med* 2016;:bjsports-2016-097017. doi:10.1136/bjsports-2016-097017

82 Newgard CD, Lewis RJ. Missing Data: How to Best Account for What Is Not Known. *JAMA* 2015;**314**:940–1. doi:10.1001/jama.2015.10516

83 El-Masri MM, Fox-Wasylyshyn SM. Missing data: an introductory conceptual overview for the novice researcher. *Can J Nurs Res Rev Can Rech En Sci Infirm* 2005;**37**:156–71.

84 Hippel PTV. How to Impute Interactions, Squares, and Other Transformed Variables. *Sociol Methodol* 2009;**39**:265–91. doi:10.1111/j.1467-9531.2009.01215.x

85 Gibbons RD, Hedeker D, DuToit S. Advances in Analysis of Longitudinal Data. *Annu Rev Clin Psychol* 2010;**6**:79–107. doi:10.1146/annurev.clinpsy.032408.153550

86 McDonald JH. *Handbook of biological statistics*. Sparky House Publishing Baltimore, MD 2009. http://biostathandbook.com/HandbookBioStatSecond.pdf

87 Therneau TM, Grambsch PM. *Modeling survival data: extending the Cox model*. New York: : Springer 2000.

88 Finch CF, Cook J. Categorising sports injuries in epidemiological studies: the subsequent injury categorisation (SIC) model to address multiple, recurrent and exacerbation of injuries. *Br J Sports Med* 2014;**48**:1276–80. doi:10.1136/bjsports-2012-091729

89 Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med* 2014;**48**:1287–93. doi:10.1136/bjsports-2011-090803

90 Shrier I, Steele RJ, Zhao M, *et al.* A multistate framework for the analysis of subsequent injury in sport (M-FASIS). *Scand J Med Sci Sports* 2016;**26**:128–39. doi:10.1111/sms.12493

91 Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965;**58**:295–300.

92 Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg Themes Epidemiol* 2005;**2**:11. doi:10.1186/1742-7622-2-11

93 Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;**95 Suppl 1**:S144-150. doi:10.2105/AJPH.2004.059204

94 Carey DL, Blanch P, Ong K-L, *et al.* Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk. *Br J Sports Med* 2016;:bjsports-2016-096309. doi:10.1136/bjsports-2016-096309

95 Rogalski B, Dawson B, Heasman J, *et al.* Training and game loads and injury risk in elite Australian footballers. *J Sci Med Sport* 2013;**16**:499–503. doi:10.1016/j.jsams.2012.12.004

96 Tan X, Shiyko MP, Li R, *et al.* A time-varying effect model for intensive longitudinal data. *Psychol Methods* 2012;**17**:61–77. doi:10.1037/a0025814

97 Hulme A, Thompson J, Nielsen RO, *et al.* Towards a complex systems approach in sports injury research: simulating running-related injury development with agent-based modelling. *Br J Sports Med* 2018;:bjsports-2017-098871. doi:10.1136/bjsports-2017-098871

98 Carey DL, Ong K-L, Whiteley R, *et al.* Predictive modelling of training loads and injury in Australian football. *ArXiv170604336 Stat* Published Online First: 14 June 2017.http://arxiv.org/abs/1706.04336

99 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. *J Sci Med Sport* Published Online First: 2017.http://www.sciencedirect.com/science/article/pii/S144024401730436X

100 Møller M, Nielsen RO, Attermann J, *et al.* Handball load and shoulder injury rate: a 31-week cohort study of 679 elite youth handball players. *Br J Sports Med* 2017;**51**:231–7. doi:10.1136/bjsports-2016-096927

101 Rossi A, Pappalardo L, Cintia P, *et al.* Effective injury prediction in professional soccer with GPS data and machine learning. *ArXiv170508079 Stat* Published Online First: 23 May 2017.http://arxiv.org/abs/1705.08079

102 Stares J, Dawson B, Peeling P, *et al.* Identifying high risk loading conditions for in-season injury in elite Australian football players. *J Sci Med Sport* Published Online First: 25 May 2017. doi:10.1016/j.jsams.2017.05.012

103 Williams S, Trewartha G, Cross MJ, *et al.* Monitoring What Matters: A Systematic Process for Selecting Training Load Measures. *Int J Sports Physiol Perform* 2016;:1–20. doi:10.1123/ijspp.2016-0337

104 Wu AD, Zumbo BD. Understanding and Using Mediators and Moderators. *Soc Indic Res* 2007;**87**:367. doi:10.1007/s11205-007-9143-1

105 Williams S, Booton T, Watson M, *et al.* Heart Rate Variability is a Moderating Factor in the Workload-Injury Relationship of Competitive CrossFit™ Athletes. *J Sports Sci Med* 2017;**16**:443–9.

106 Sampson JA, Murray A, Williams S, *et al.* Injury risk-workload associations in NCAA American college football. *J Sci Med Sport* Published Online First: 22 May 2018. doi:10.1016/j.jsams.2018.05.019

107 Garcia TP, Marder K. Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model. *Curr Neurol Neurosci Rep* 2017;**17**:14. doi:10.1007/s11910-017-0723-4

108 Yang J, Zaitlen NA, Goddard ME, *et al.* Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 2014;**46**:100–6. doi:10.1038/ng.2876

109 Russell MA, Almeida DM, Maggs JL. Stressor-related drinking and future alcohol problems among university students. *Psychol Addict Behav* 2017;**31**:676–87. doi:10.1037/adb0000303

110 Bolger N, Laurenceau J-P. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. Guilford Press 2013.

111 Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. New York: : Cambridge University Press 2007. http://www.loc.gov/catdir/enhancements/fy0668/2006040566-t.html (accessed 12 Apr 2016).

112 Horton NJ, Switzer SS. Statistical Methods in the Journal. *N Engl J Med* 2005;**353**:1977–9. doi:10.1056/NEJM200511033531823

113 Nielsen RØ, Malisoux L, Møller M, *et al.* Shedding Light on the Etiology of Sports Injuries: A Look Behind the Scenes of Time-to-Event Analyses. *J Orthop Sports Phys Ther* 2016;**46**:300–11. doi:10.2519/jospt.2016.6510

114 Wu L, Liu W, Yi GY, *et al.* Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues. J. Probab. Stat. 2012. doi:10.1155/2012/640153

115 Andersen PK, Syriopoulou E, Parner ET. Causal inference in survival analysis using pseudo-observations. *Stat Med* 2017;**36**:2669–81. doi:10.1002/sim.7297

116 Hansen SN, Andersen PK, Parner ET. Events per variable for risk differences and relative risks using pseudo-observations. *Lifetime Data Anal N Y* 2014;**20**:584–98. doi:http://dx.doi.org.ezproxy.library.ubc.ca/10.1007/s10985-013-9290-4

117 Hickey GL, Philipson P, Jorgensen A, *et al.* Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol* 2016;**16**. doi:10.1186/s12874-016-0212-5

118 Mahmood A, Ullah S, Finch CF. Application of Survival Models in Sports Injury Prevention Research: A Systematic Review. *Br J Sports Med* 2014;**48**:630–630. doi:10.1136/bjsports-2014-093494.190

119 Carey DL, Crossley KM, Whiteley R, *et al.* Modelling Training Loads and Injuries: The Dangers of Discretization. *Med Sci Sports Exerc* Published Online First: June 2018. doi:10.1249/MSS.0000000000001685

120 Maldonado G, Greenland S. The importance of critically interpreting simulation studies. *Epidemiology* 1997;:453–456.

121 Trail JB, Collins LM, Rivera DE, *et al.* Functional data analysis for dynamical system identification of behavioral processes. *Psychol Methods* 2014;**19**:175–87. doi:10.1037/a0034035

122 de Haan-Rietdijk S, Kuppens P, Hamaker EL. What's in a Day? A Guide to Decomposing the Variance in Intensive Longitudinal Data. *Front Psychol* 2016;**7**. doi:10.3389/fpsyg.2016.00891

123 Biddle SJH, Edwardson CL, Wilmot EG, *et al.* A Randomised Controlled Trial to Reduce Sedentary Time in Young Adults at Risk of Type 2 Diabetes Mellitus: Project STAND (Sedentary Time ANd Diabetes). *PLOS ONE* 2015;**10**:e0143398. doi:10.1371/journal.pone.0143398

124 Casals M, Finch CF. Sports Biostatistician: a critical member of all sports science and medicine teams for injury prevention. *Inj Prev* 2016;:injuryprev-2016-042211. doi:10.1136/injuryprev-2016-042211

Figure 1 – The workload—injury aetiology model. Key features include the multifactorial nature of injury, between- and within- athlete differences in risk, and a recursive loop.

Figure 2 – Complex systems model of athletic injury. Web of determinants are shown for an ACL injury in basketball players (A), and in a ballet dancer (B)

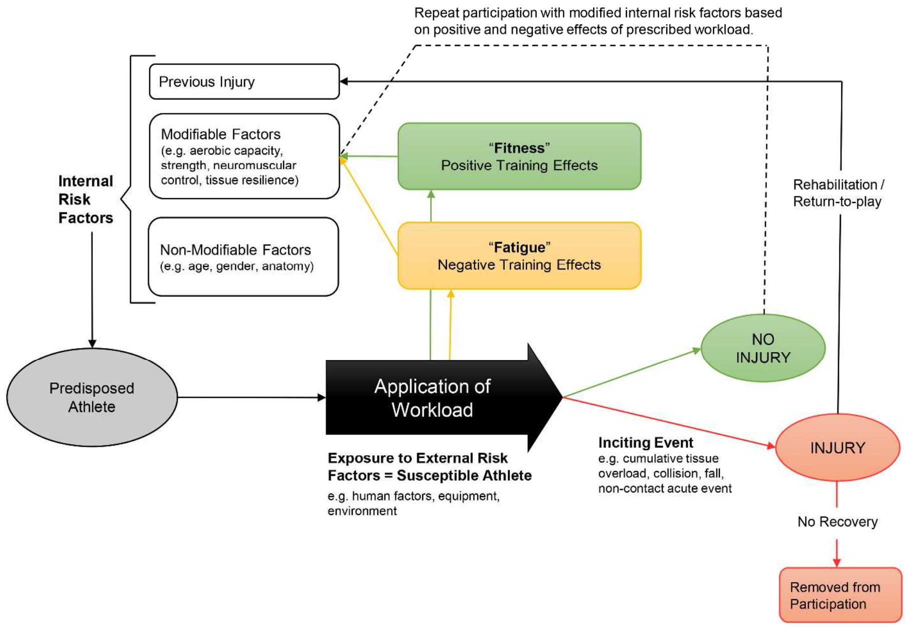

Figure 1 – The workload—injury aetiology model. Key features include the multifactorial nature of injury, between- and within- athlete differences in risk, and a recursive loop.

299x199mm (300 x 300 DPI)

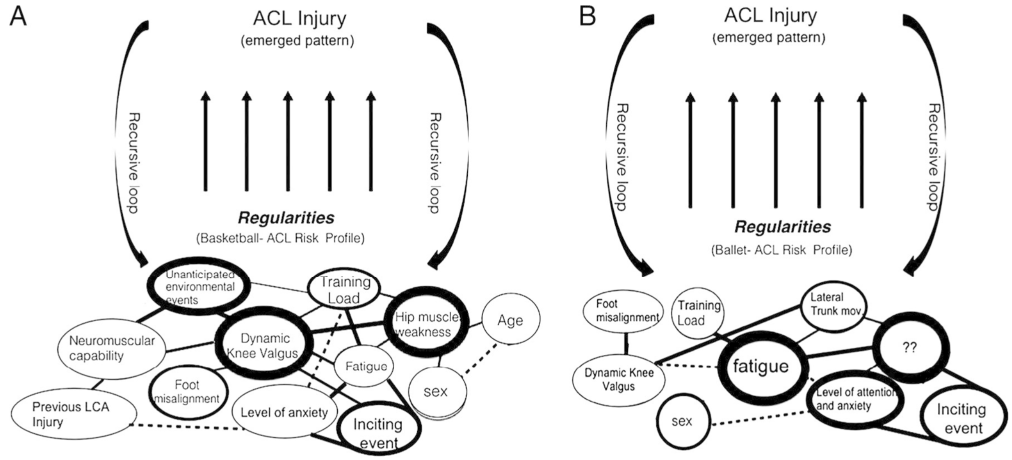

Figure 2 – Complex systems model of athletic injury. Web of determinants are shown for an ACL injury in basketball players (A), and in a ballet dancer (B)

152x69mm (300 x 300 DPI)

APPENDIX 1 – Systematic Search Strategy and Article Selection

Table A1 – Example search categories and terms:

Category	Search Terms
Population	Athlet*, sport*
Method	Systematic review, Consensus statement
Outcome (Injury)	Injur*, illness*, strain, sprain, incidence, overuse, overreach*, accidents, stress, wellness, recover*
Workload	Training, resistance training, external load, internal load, workload, acute:chronic workload ratio, congested calendar, physical exertion, session RPE, global position systems, accelerometry, intensity, duration, physical fitness, fatigue

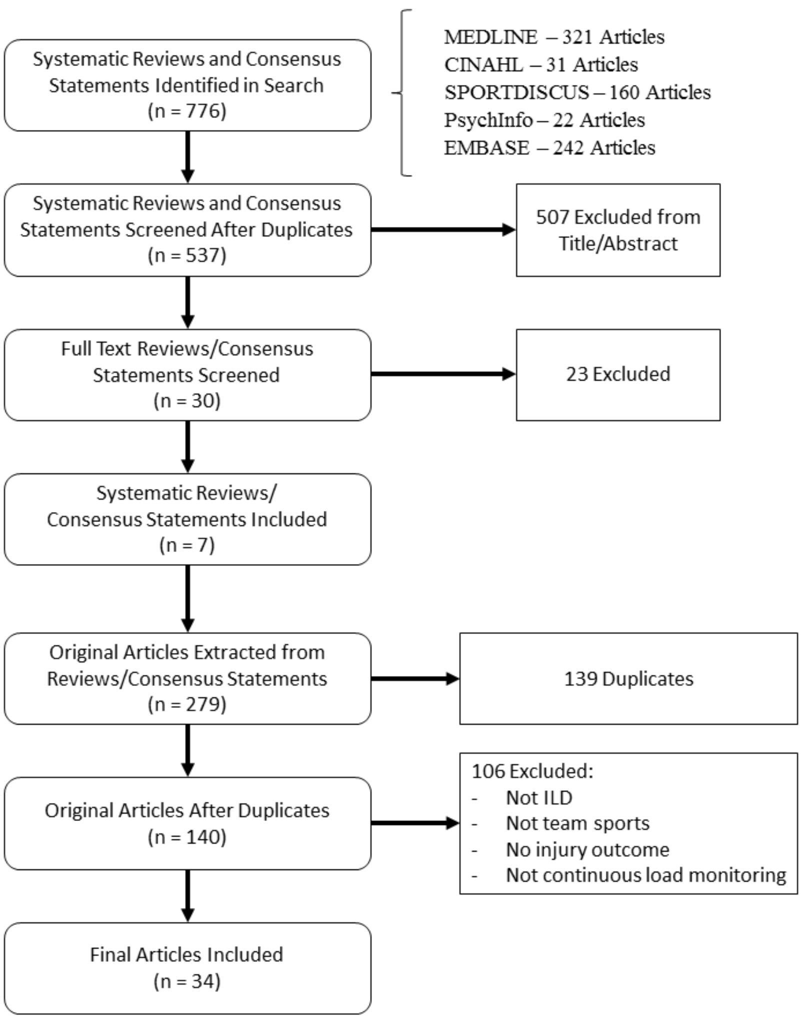

Figure A1 – Flow chart of included articles

APPENDIX 2 – TABLE A2: Expanded Table of 3-Fold Alignment (With Descriptions)

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Correlation (Pearson and Spearman)	10	X	X	X	X	X	X	X	7 of the 10 articles correlated team loads with injury incidence (team-level). Of those at the individual level, 2 looked at # of pre-season sessions and % in-season completed, and 1 looked at training load and injury subscale on the REST-Q.
Anderson - 2003 Bresciani - 2010 Books - 2008 Gabbett - 2004 Gabbett Jenkins - 2011 Killen 2010 Mallo Dellal - 2012 Murray Gabbett - 2016 Owen - 2015 Windt – 2016		Can only handle one x and one y variable.	Correlation could be at the team level (training load and # of injuries), or on individual level with quantitative outcome, but cannot differentiate within/between athlete differences.	No interactions between multiple predictors	No, assumes independent observations.	Assumes one observation per research participant/study unit, so participants couldn't have different numbers of observations.	In this case the correlation assumes independent observations so dependency is not taken into account.	Could, through having time as one of the variables, but cannot account for temporality	
Unpaired t-test	6	X	X	X	X	X	X	X	Generally, compared injured and uninjured players across a season. A group comparison of loading across the year fails to account for between/within athlete considerations and doesn't specify temporality.
Dennis et al., 2003 Dennis et al., 2005 Duhig et al - 2016 Owen et al., 2015 Saw et al., 2011 Visnes Bahr., 2013		Can only handle one x (grouping) and one y (outcome) variable	Only between-athlete differences (injured vs. uninjured)	No interactions between multiple predictors	Independent samples t-test compare group means on either a time-varying (e.g average workload) or time-invariant variable (e.g. height), not both.	No, assumes one observation per research participant/study unit, so by design forces a balanced set (1 observation per participant)	By definition, assumes independence	None included time in the analysis.	

		Themes of theoretical model			**Themes of temporal design: intensive longitudinal data**				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Chi-Square Tests	1	X	X	X	X	X	X	X	
Murray - 2016 - IJSPP		Only examines load groups and injury incidence	Examines differences in injury incidence across different load groups only	No interactions between multiple predictors	Included load groups and injury incidence only	Forces 1 observation (aggregated variable) per participant	Designed for independent observations and groups	Only included load groups and injury incidence	
Relative Risk Calculations	8	O	X	X	X	X	X	X	
Bowen - 2016 Dennis - 2003 Dennis - 2005 Hulin - 2014 Hulin - 2016 Hulin - 2016 Murray - 2016 - Scand.		In some cases, authors examined relative risks of loading groups after subdividing across another variable, like chronic workload, making it multifactorial. Other authors only examined risks across load groups.	No differentiation, and independence assumed	No interactions between multiple predictors	Only loading (time-varying variables included)	Assumes independence of observations	Assumes independence	Uses weeks as unit of analysis but no incorporation of time into the calculations	Many of these RR approaches seem to use RRs that traditionally require independence, but do not account for this in their analysis.

Method	n	Themes of theoretical model: Multifactorial aetiology	Themes of theoretical model: Between and within-athlete differences	Themes of theoretical model: Complex system	Themes of temporal design: intensive longitudinal data: Includes time-varying and time-invariant variables	Themes of temporal design: intensive longitudinal data: Missing/unbalanced data (due to rep. measures)	Themes of temporal design: intensive longitudinal data: Repeated measure dependency	Themes of temporal design: intensive longitudinal data: Incorporates time in the analysis	Statistical summary and typical uses
Regression (logistic, linear, multinomial)	13	O	X	X	X	X	X	X	
Arnason 2004 - Regular logistic Bowen - 2016 - Regular logistic Brink - 2010 - Regular multinomial Colby - 2014 - Regular logistic Duhig - 2016 - Regular logistic Gabbett Domrow - 2007 - Regular linear Hulin - 2014 - Regular logistic Hulin - 2016 - Regular logistic Hulin - 2016 - Regular logistic Murray - 2016 - Regular logistic (Scand) Owen - 2015 - Regular linear Rogalski - 2013 - Regular logistic Visnes Bahr - 2013 - Regular logistic		Some authors include multiple variables, others only single load measurements independently	Assume independent observations, so cannot examine within-athlete differences	None included interactions between predictors	Assumes 1 observation per unit Becomes time invariant on aggregation In this case, some authors have included both, but in doing so violate the independence assumption	No, assumes one observation per athlete or research unit	Assumes independence. Visnes & Bahr (2013) do take it into account through logistic regression, but do not have the benefit of ILD	Assumes 1 observation per unit	
Paired t-test	2	X	X	X	X	X	√	√	
Saw et al., 2011 Dennis et al., 2003		No, only one variable (before/after) and outcome (load)	Only examines within-athlete differences	No interactions between multiple predictors	No - only time-varying variables	All subjects must have 2 'observations'	By definition, accounts for repeated measures through paired sample	Time is incorporated as pre-injury and 'injury' blocks of time	

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Repeated measures ANOVA (One- or two-way ANOVA allowing for between- and within)	5	O	O	X	O	X	√	√	Killen used 1-way ANOVA to compare the load in early pre-season compared to late pre-season, and used chi-square analyses to compare injury rate in the early and late pre-season. Uses the two separate analyses to tentatively link load leads to injury. Gabbett, 2004 performed a 2-way ANOVA comparing loads (season X month), so time is included.
Ehrman - 2016 Gabbett - 2004 Malisoux - 2013 Murray - 2016 Killen - 2011		Murray (2016) compared part of season by training load group, Malisoux (2013) and Ehrmann (2016) only compared injury and pre-injury blocks	In some cases (Murray, 2016), a two-way repeated measures ANOVA can examine between and within athlete differences in risk	No interactions between multiple predictors	Murray (2013) compared load group by season period	Assume sphericity	Yes, by definition	Yes, as season period, or as pre-injury and injury period	
Cox proportional hazards model	1	√	X	X	X	√	√	√	
Malisoux - 2013		Included volume and intensity of training along with age and sex	Cox PH conducted at the team/school level examined between-athlete differences	No interactions between factors	Only included average weekly load and average intensity	Can handle unbalanced data	By using time-to-event as the outcome, Cox-PH robust to this dependency	Uses time to event in analysis	

		Themes of theoretical model			Themes of temporal design: intensive longitudinal data				Statistical summary and typical uses
Method	n	Multifactorial aetiology	Between and within-athlete differences	Complex system	Includes time-varying and time-invariant variables	Missing/unbalanced data (due to rep. measures)	Repeated measure dependency	Incorporates time in the analysis	
Generalised estimating equations (modeled through logistic and poisson regression)	6	O	X	X	O	√	√	O	Most GEE approaches accounted for repeated measures, but Clausen et al (2014) used them to cluster players within teams), averaging exposure throughout the entire season.
Clausen - 2014 Cross - 2016 Dennis - 2004 Gabbett - 2010 Gabbett Domrow - 2007 Veugeleurs - 2016		Most authors used multiple variables with GEEs, although some only used GEE to account for repeated measures (Dennis, 2004, Gabbett Domrow, 2007)	Provides an 'average' effect for all athletes, but controls for the clustering	No complex interactions between factors	Some authors only predicted injury (y/n) based on workload variables	GEEs handle unbalanced data well	GEE accounts for clustering	Do not incorporate time explicitly into the modelling process, and none of the authors included time in the model.	
Multilevel Modeling	1	√	√	X	√	√	√	X	
Windt et al - 2016		Multiple physical outputs included and pre-season training	Within-athlete risks determined from level 1 variables, between-athlete from level 2	No complex interactions between factors	Pre-season and player variables along with training variables	Multilevel models are robust to missing/unbalanced data	Via random effects for each player	Analysed weekly risk and subsequent week risk; did not directly incorporate time	
Frailty model	1	√	√	X	√	√	√	√	Included a many physical output metrics that were dichotomised into high/low categories and calculated RRs. Multicollinearity was not considered
Gabbett Ullah - 2012		Yes, previous injury and physical outputs	Unclear (I think yes due to the frailty term)	No complex interactions between factors	Yes (injury history)	Yes (robust to unbalanced data)	Yes (frailty model allows for dependency of recurrent events)	Yes, since this is a time-to-event analysis.	