wordmaker

# Introduction

Over the last four decades sport related injuries have received increased research attention [@Ivarsson2017]. This attention is unsurprising given the high incidence [@Rosa2014; @Sheu2016], and undesirable physical and psychological effects of sports injuries [@Leddy1994; @Brewer2012]. To mitigate against both the increasing incidence and undesirable consequences of injury, research has identified several psychological [@Slimani2018], anatomical [@Murphy2003], biomechanical [@Neely1998; @Hughes2014] and environmental [@Meeuwisse2007] factors associated with sports injury occurrence. Indeed, several models of injury causation have been proposed that highlight the multifactorial nature of injury occurrence [@Kumar2001; @Meeuwisse2007; @Wiese-Bjornstal2009], of which one of the most widely cited was developed by Williams and Anderson [Fig \ref{fig:fig1}; @Andersen1988; @Williams1998].

Williams and Andersen’s [@Williams1998] model proposed that when faced with a potentially stressful athletic situation, an athlete’s personality traits (e.g., hardiness, locus of control and competitive trait anxiety), history of stressors (e.g., major life events and previous injuries) and coping resources (e.g., general coping behaviours) will contribute to their response, either interactively or in isolation. Central to the model is the stress response, which reflects the bi-directional relationship between athletes’ appraisal of, and response to, a stressful athletic situation. The model predicts that athletes who have a history of many stressors, personality traits that intensify the stress response and few coping resources, will exhibit greater attentional (e.g., peripheral narrowing) and/or physiological (e.g., increased muscle tension) responses that put these individuals at greater risk of injury.

Within Williams and Andersen’s [@Williams1998] model, major life events, a component of an athlete’s history of stressors, most consistently predicts injury occurrence [@Williams2007]; specifically, major life events with a negative, as opposed to positive, valence [@Passer1983a; @Maddison2005]. However, personality traits and coping resources have also been found to predict injury, with for example, athletes more likely to sustain an injury if they have poor social support and psychological coping skills, and high trait anxiety and elevated competitive state anxiety; compared to athletes with the opposite profile. [@Smith1990; @Lavallee1996; @Ivarsson2010]. However, the amount of variance explained by the psychosocial factors proposed by the model has been modest, typically between 5 - 30% [@Galambos2005; @Ivarsson2010]; suggesting other factors are also likely to contribute to injury occurrence.

While the psychosocial factors proposed in Williams and Andersen’s [@Williams1998] model have received the most research attention, the mechanisms through which these factors are proposed to exert their effect have remained under-investigated in the literature. To elaborate, the model suggests that injuries are likely to occur through either increased physiological arousal resulting in increased muscle tension and reduced flexibility or attentional deficits caused by increased distractibility and peripheral narrowing. However, to date, the research has largely focused on attentional deficits [@Andersen1999; @Rogers2005; @Wilkerson2012a; @Swanik2007]. For example, Andersen and Williams [@Andersen1999] measured peripheral and central vision during high and low stress conditions and found athletes with high life event stress coupled with low social support had greater peripheral narrowing under stressful conditions compared to athletes with the opposing profile; these athletes went on to sustain an increased number of injuries during the following season. Indeed, Rodgers and Landers [@Rogers2005] supported Andersen and Williams’s [@Andersen1999] earlier findings reporting that peripheral narrowing under stress mediated 8.1% of the relationship between negative life events and injury. However, the remaining variance between negative life events and athletic injury through the other proposed mechanisms, such as increased muscle tension and reduced motor control, remains to be explored [cf. @Williams1998].

One challenge faced by researchers addressing the sports injury problem with a psychological lens is the multifactorial nature of injury, and the possible contribution of other non-psychological factors to the stress response [@Meeuwisse2007; @Wiese-Bjornstal2009]. For example, a large body of research indicates that training-related stress is also likely to be related to the stress response and injury occurrence [@Lee2017; @Djaoui2017], and may account for the unexplained variance from the psychological predictors of injury. Appaneal and Perna [@Appaneal2014] proposed the biopsychosocial model of stress athletic injury and health (BMSAIH) to serve as an extension to Williams and Andersen’s [@Williams1998] model and to address some of these issues. To elaborate, the BMSAIH aimed to clarify the mediating pathways between the stress response and injury, consider other health outcomes and behavioural factors that impact sports participation, and integrate the impact of training on athletes’ health [@Appaneal2014]. The central tenet of the BMSAIH is that psychosocial distress (e.g., negative life events) may act synergistically with training-related stress as a result of high-intensity and high-volume sports training, and “widen the window of susceptibility” [@Appaneal2014, p. 74] to a range of undesirable health outcomes including illness and injury. Consequently, the BMSAIH provides a framework for future research to build on Williams and Andersen’s [@Williams1998] model, by including other physiological markers of training-related stress, which together may provide greater insight into the injury process.

Although research supporting the BMSAIH has mainly focused on the relationship between hormonal responses to training and injury occurrence [@Perna1995; @Perna1997; @Perna2003], other research has identified additional markers of training-related stress that are associated with an increased risk of injury; for example, heart rate variability [@Bellenger2016; @Williams2017], postural stability [@Romero-Franco2014] and muscle stiffness [@Pruyn2015]. However, these markers are often studied in isolation without an assessment of the psychosocial factors that are known to contribute to injury, thereby limiting our understanding of how psychosocially and physiologically derived stress may contribute synergistically to injury occurrence. Recently, Bittencourt et al. [@Bittencourt2016] suggested that to better understand the multifactorial nature of sports injuries, research needs to move away from studying risk factors in isolation and instead adopt a complex systems approach to injury. Such an approach posits that injury may arise from a complex “web of determinants” [@Bittencourt2016, p. 3], where different factors interact in unpredictable and unplanned ways, but result in a global outcome pattern of either adaptation or injury.

A challenge when adopting a complex systems approach is using an appropriate analysis technique that is able to capture the uncertainty and complexity of the relationships between different variables. One technique that provides a solution to this challenge is Bayesian network (BN) modelling. BN’s allow the construction of graphical probabilistic models using the underlying structure that connects different variables [@Scutari2014]. The learned structure can be used for inference by obtaining the posterior probabilities of a particular variable for a given query (e.g., if the value of variable A is x and the value of variable B is y, what is the probability variable C being value z?). Furthermore, BN’s do not distinguish between dependent and independent variables as they are a form of unsupervised learning, which is a strength over regression or structure equation models when the underlying relationship in the network may not be known [@Olmedilla2018].

To summarise, despite offering a possible framework to build on the research stemming from Williams and Andersen’s [@Williams1998] model, there remains and opportunity to explore other physiological stress-related markers proposed by the BMSAIH, in addition to the already well-established psychological characteristics known to be related to injury [@Appaneal2014]. Furthermore, research has typically not captured changes in both psychosocial factors and stress-related physiological markers that may occur between initial measurement and injury occurrence. Given the exploratory nature of the current study the following objectives were defined:

* Identify suitable markers of stress that can be easily captured in a large cohort of athletes in a field based setting.
* Capture the markers of stress and injury occurrence using a prospective, repeated measures design.
* Explore and evaluate the relationships between the markers of stress and injury using Bayesian network modelling.

# Methods

## Participants

A total of 351 athletes (male: *n* = 231, female: *n* = 120) were initially recruited for the study. Participants had an average age of 22.0 7.0 years and represented a range of team (football, rugby, netball, cricket, lacrosse, basketball and field hockey) and individual sports (athletics, tennis, weightlifting, gymnastics, judo, swimming and golf from a British University and local sports clubs (Table ). Participants self-rated competitive level ranged from recreational to international standard. A total of 126 (49.03%) participants had sustained an injury in the 12 months prior to the start of the study (male: *n* = 114 [49%], female; *n* = 48 [40%]). At the start of the study, all participants were injury free (no modifications to their usual training routine due to a sport related medical problem for a minimum of four weeks). Participants were engaged in training for their respective sports for at least five hours per week. Ethical approval was obtained from the University ethics committee prior to the start of the study and all participants provided informed consent.

## Measures

### Major life events

A modified version of the Life Events Survey for Collegiate Athletes (LESCA) was used to measure participants’ history of life event stress [@Petrie1992]. Modifications were made to the LESCA to ensure the suitability of the items for the study population (S1 Table). The LESCA comprises 69 items that reflect possible life events that participants may have experienced. Example items include, “Major change in the frequency (increased or decreased) of social activities due to participation in sport”, “Major change in the amount (more or less) of academic activity (homework, class time, etc)” and “Major change in level of athletic performance in actual competition (better or worse)”. Participants were asked to rate the perceived impact of each life event they had experienced within the last 12 months on an 8-point Likert scale anchored at -4 () and +4 (). Negative and positive life event scores were calculated by summing the negative and positive scores respectively. A score for total life events was calculated by summing the absolute values for both negative and positive events. Petrie (1992) [@Petrie1992] reported test-retest reliabilities at 1-week and 8-weeks with values ranging from .76 to .84 (p < .001) and .48 to .72 (p < .001) respectively. Petrie also provided evidence of discriminant, convergent and predictive validity. The LESCA is the most widely used measure of major life events for athletes in the sports injury literature. For this study, Composite Reliability [@Fornell1981] was preferred to Cronbach’s alpha as it does not assume parallelity (i.e., all factor loadings are constrained to be equal, and all error variances are constrained to be equal) and instead takes into consideration the varying factor loadings of the items in the questionnaire. The composite reliability for the LESCA in this study was 0.84.

### Reinforcement Sensitivity Theory Personality Questionnaire

A revised version of the Reinforcement Sensitivity Theory Personality Questionnaire (RST-PQ) was used to measure motivation, emotion, personality and their relevance to psychopathology [@Corr2016c]. The revised version of the RST-PQ comprises 51 statements that measure three major systems: Fight-Flight-Freeze System (FFFS; e.g., “I am the sort of person who easily freezes-up when scared”), Behavioural Inhibition System (BIS; e.g., “When trying to make a decision, I find myself constantly chewing it over”) and four Behavioural Approach System (BAS) factors; Reward Interest (e.g., “I regularly try new activities just to see if I enjoy them”), Goal Drive Persistence (e.g., “I am very persistent in achieving my goals”), Reward Reactivity (e.g., “I get a special thrill when I am praised for something I’ve done well”) and Impulsivity (e.g., “I find myself doing things on the spur of the moment”). Participants rated each item on a scale from 1 () to 4 () to reflect how well each statement described their personality in general. The responses to items associated with each subscale (FFFS, BIS, RI, GDP, RR and I) were summed to give a total score that was subsequently used for further analysis. The composite reliabilities for each subscale were; BIS = 0.92, FFFS = 0.77, GDP = 0.87, I = 0.71, RI = 0.77, RR = 0.81. Further details regarding the revised RST are in S1 Appendix.

### Heart rate variability

A Polar V800 heart rate monitor (HRM) and Polar H7 Bluetooth chest strap (Polar OY, Finland) was used to collect inter-beat interval (IBI) data. IBI recordings using the Polar V800 are highly comparable (ICC = 1.00) with ECG recordings [@Giles2016], which are considered the gold standard for assessing HRV. In addition, HRV indices calculated from IBI and ECG data have shown a strong correlation (r = .99) in athletes [@Caminal2018] and under spontaneous breathing conditions [@Plews2017].

### Musculoskeletal properties

A handheld myometer (MyotonPRO, Myoton AS, Tallinn, Estonia) was used to measure muscle stiffness. The MyotonPRO is a non-invasive, handheld device that applies a mechanical impulse of 0.40 N for 0.15 ms perpendicular to the surface of the skin. The impulse causes natural damped oscillations in the tissue, which are recorded by a three-axis digital accelerometer sensor in the device. The raw oscillation signal is then processed, and the stiffness parameter is calculated [@Agyapong-Badu2016]. The MyotonPRO has previously been reported to be a reliable and valid tool for the measurement of in-vivo tissue stiffness properties [@Chuang2013; @Pruyn2016; @Nair2014], and has demonstrated good internal consistency (coefficient of variation < 1.4%) over sets of 10 repetitions [@Aird2012].

### Postural stability

Postural stability was assessed with a modified version of the balance error scoring system (mBESS) based on the protocol recommended by @Hunt2009. In total, each trial of the mBESS was performed without shoes [@McCrory2013] and included six stances in the following order; dominant leg (DL; standing on the dominant foot with the non-dominant foot at approximately 30-degrees of hip flexion and 45-degrees of knee flexion), non-dominant leg (NDL; standing on the non-dominant foot with the dominant foot at approximately 30-degrees of hip flexion and 45-degrees of knee flexion) and tandem leg stance (TS; standing heel-to-toe with the non-dominant foot behind the dominant) on firm and foam (Alcan airex AG, Sins, Switzerland) surfaces respectively (Fig ). To determine leg dominance, participants were asked their preferred leg to kick a ball to a target, and the chosen limb was labelled as dominant [cf. @VanCingel2017]. Participants were asked to maintain each stance for a total of 20-seconds. Participants hands were placed on hips at the level of the iliac crests. A Sony DSC-RX10 video camera (Sony Europe Limited, Surrey, United Kingdom) was used to record each participants performance during the mBESS.

The error identification criteria from the original BESS protocol was used by the lead researcher who scored all the BESS trials. One error was recorded if any of the following movements were observed during each trial: a) lifting hands off iliac crests; b) opening eyes; c) stepping, stumbling, or falling; d) moving the thigh into more than 30 degrees of flexion or abduction; e) lifting the forefoot or heel; and f) remaining out of the testing position for more than 5-seconds [@Riemann1999d]. A maximum score of 10 errors was possible for each stance. Multiple errors occurring simultaneously were recorded as one error. A participant was given the maximum score of 10 if they remained out of the stance position for more than 5-seconds. To calculate limb asymmetry, the DL and NDL leg score was calculated by summing the DL and NDL errors respectively. A total score was calculated by summing the total number of errors recorded on all stances (DL, NLD and TS, on foam and firm surfaces). To assess the intra-rater reliability, a single measurement, absolute agreement, two-way mixed effects model for the intraclass correlation [ICC; @Koo2016] was used on a sample of 40 participants from the first time point. The test-retest scoring of BESS resulted in a “good” to “excellent” ICC score (ICC = 0.93, 95% confidence interval = 0.88 - 0.96), indicating the scoring was reliable.

### Injury

Participants self-reported any injuries they sustained at each data collection during the study period. An injury was defined as any sports related medical problem causing the athlete to miss or modify their usual training routine for at least 24 hours [@Fuller2006; @Fuller2007b; @Timpka2014]. Minor scrapes and bruises that may require certain modifications (e.g., strapping or taping) but did not limit continued participation were not considered injuries [cf. @Appaneal2009]. Injury status (did / did not sustain an injury) served as the main outcome measure.

## Procedure

At the start of the UK academic year (September 2016 and 2017), coaches of sports teams at a British University and local sports clubs were contacted and informed about the study. With the coaches’ permission, the lead researcher attended training sessions to inform the athletes about the overall purpose of the study and the requirements of participation. To be eligible for the study athletes had to be injury free (no modifications to their usual training routine due to a sport related medical problem for a minimum of four weeks) and training a minimum of five hours per week. Athletes who met the criteria and volunteered to take part in the study were invited to attend scheduled testing sessions. A repeated measures prospective cohort design was used to assess athletes’ major life events, stress-related physiological markers and injury status over two consecutive twelve-month periods between September 2016 and September 2018. Each participant was asked to attend a total of 3 data collections over a twelve-month period, with each data collection separated by a four-month interval (Fig ). Participants provided informed consent before data collection commenced.

For the first three data collections (T1, T2 and T3), participants followed the same protocol in a specific order (Fig ). To ensure all measures could be collected within a viable time-frame, participants were separated into two groups. The first group completed all computer-based measures followed by all physical measurements, whereas the second group completed all physical measurements followed by computer-based measures. Participants were randomly assigned to one of the two groups and remained in those groups across all time points.

### Questionnaires

The questionnaires, which included demographic information, the LESCA, RST-PQ (T1, T2, T3) and injury status (T2, T3, T4) were completed on-line (SurveyMonkey Inc., USA, www.surveymonkey.com). The instructions for the LESCA were modified at T2 and T3 so that participants reported major life events that had occurred since the previous testing session. For injury reporting, participants were asked to record any injuries that they had sustained since the last data collection. The data were downloaded from surverymonkey.com and imported into R [@RCoreTeam2019] for analysis purposes.

### HRV

To minimise potential distractions, participants were directed to a designated quiet area in the laboratory where IBI data were recorded. Participants were instructed to turn off their mobile devices to avoid any interference with the Bluetooth sensor. Each chest strap was dampened with water and adjusted so it fitted tightly but comfortably, as outlined by Polar’s guidelines. Participants were seated and asked to remain as still as possible for the duration of the recording. No attempt was made to control participants respiratory frequency or tidal volume [@Denver2007]. Inter-beat interval (IBI) data was collected for 10-minutes at a sampling frequency of 1000 Hz.

Raw, unfiltered IBI recordings were exported from the Polar Flow web service as a space delimited .txt file and imported into R [@RCoreTeam2019] where the *RHRV* package [@Rodriguez-Linares2017] was used to calculate HRV indices. Raw IBI data was filtered using an adaptive threshold filter, and the first 3-minutes and last 2-minutes of each recording were discarded, leaving a 5-minutes window that was used to calculate the root mean square of successive differences (RMSSD) in RR intervals following the recommendations for short term IBI recordings [@Laborde2017; @Malik1996]. RMSSD was calculated as:

Where N is the length of the time series, and the RR interval between beats and , where each beat position corresponds to the beat detection instant.

### Muscle stiffness

To assess muscle stiffness, participants lay horizontally on a massage bed and four testing sites were identified on each lower limb. The muscle belly of the rectus femoris (RF), biceps femoris (BF), medial gastrocnemius (MG) and lateral gastrocnemius (LG) sites were identified using a visual-palpatory technique to determine the exact location of each site [@Chuang2012]. The visual-palpatory technique required the participant to contract the target muscle to aid the lead researcher to visually identify the muscle. The participant was then asked to relax the muscle and the muscle was palpated to locate the muscle belly. A skin safe pen (Viscot all skin marker pen, Viscot Medical LLC, NJ) was used to mark the testing site in the centre of the muscle belly.

After the eight testing sites had been identified, the testing end of the MyotonPRO (diameter = 3 mm) was positioned perpendicular to the skin on the testing site. A constant pre-load of 0.18 N was applied for initial compression of subcutaneous tissues. The device was programmed to deliver five consecutive impulses, separated by a one second interval [@Morgan2018]. For each impulse, the device computed stiffness values, with the median of the five values being saved by the device for further analysis. In accordance with Myoton.com, a set of five measurements with a coefficient of variation (CV) of less than 3% was accepted. Sets of measurements above 3 % were measured again to ensure the reliability of the data. The CV was calculated in real time by the device after each set of measurements. Measurements saved on the device were uploaded to a computer using MyotonPRO software and imported in R [@RCoreTeam2019] for further analysis. For each participant, the sum of all eight testing sites was calculated to provide a total lower extremity stiffness score and was used for further analysis.

### Postural stability

Instructions for the mBESS were read to each participant and a demonstration of the positions was provided by the research assistant. For each position, participants were instructed to close their eyes, rest their hands on their iliac crests and remain as still as possible for 20-seconds. Participants were instructed to return to the testing position as quickly as possible if they lost their balance. The video recording was started prior to the first stance position and stopped after all stances had been completed. Each completed mBESS protocol took approximately 4 minutes. Only one trial was performed to avoid familiarisation effects across the repeated measurement [cf. @Valovich2003]. The video recordings for each participant were imported from the recording equipment (Sony DSC-RX10) and the lead researcher scored each trial using the error identification criteria.

## Data Analysis

Two Bayesian Networks (BN) were used to explore the relationships between the psychological measures, physiological markers of stress and sports injury. A BN is a graphical representation of a joint probability distribution among a set of random variables, and provides a statistical model describing the dependencies and conditional independences from empirical data in a visually appealing way [@Scutari2014]. A BN consists of arcs and nodes that together are formally known as a directed acyclic graph (DAG), where a node is termed a parent of a child if there is an arc directed from the former to the latter [Fig \ref{fig:fig5}; @Pearl1988]. However, the direction of the arc does not necessarily imply causation, and the relationship between variables are often described as probabilistic instead of casual [@Scutari2014]. The information within a node can be either continuous or discrete, and a complete network can contain both continuous and discrete nodes; however, discrete networks are the most commonly used form of BN [@Chen2012]. In discrete networks, conditional probabilities for each child node are allocated for each combination of the possible states in their parent nodes and can be used to assess the strength of a dependency in the network.

In order to use discrete networks, continuous variables must first be split into categorical levels. When there are a large number of variables in the network, limiting the number of levels has the benefit of producing a network that is more parsimonious in terms of parameters. For example, a network with 10 variables each with two levels has 100 (10^2) possible parameter combinations, however the same network with three levels has 1000 (10^3) possible parameter combinations, the latter being significantly more computationally expensive. Using a larger number of splits in the data also comes at a cost of reducing the statistical power in detecting probabilistic associations, and reduces the precision of parameter estimates for the probabilistic associations that are detected because it reduces the sample-size-to-parameters ratio [@Scutari2014]. Typically, no more than three levels have been used in Bayesian networks in the sports injury literature [@Olmedilla2018; @Ruiz-Perez2019]

Learning the structure of the network is an important step in BN modelling. The structure of a network can be constructed using expert knowledge and/or data-driven algorithm techniques (e.g., search and score, such as hill climbing and gradient descent algorithms; Scutari & Denis, 2014). The learned structure can then be used for inference by querying the network[[1]](#footnote-1) and obtaining the posterior probabilities of a particular node for a given query. The posterior distribution can be obtained by , where the learned network with structure and parameters , are investigated with new evidence using the information in [@Scutari2014]. In the example network presented in Fig , new values assigned to each of the parent nodes (e.g., both set to “Low”) could be used to investigate what effect the new information has on the state of the child node (conditional probability of a particular state of the child node). In a more complex network containing many nodes, the outcome of a particular node can be assessed conditional on the states of any subset of nodes in the network. BNs therefore provide a unique and versatile approach to modelling a set of variables to uncover dependency structures within the data.

BNs have recently been used in the sport psychology literature [@Olmedilla2018; @Fuster-Parra2017] and offer several benefits over traditional statistical analysis. For example, predictions can be made about any variable in the network, rather than there being a distinction between dependent and independent variables in the data, such as in linear regression models that are often used within the sport psychology literature [@Olmedilla2018; @Bittencourt2016]. Furthermore, the structure of a network can be obtained from both empirical data *and* prior knowledge about the area of study; the latter being particularly useful when there are a large number of variables in the network, or only a small number of observations are available in the data [@Xiao-xuan2007]. In such instances, a purely data driven approach to learning the network would be time-consuming due to the large parameter space, and inefficient at identifying an approximation of the true network structure. Prior knowledge about dependencies between variables can therefore be included in the network structure, while still allowing a data driven approach for unknown dependencies, to improve the overall computation of the network structure [@Heckerman1995; @Xu2015]. The following sections detail the steps taken in the current study to firstly prepare the data for each network, and then obtain the structure of each network that was used for further inference.

### First network

#### Data Preparation

Of the 351 participants that were initially recruited for the study, 94 only completed the first time point, and therefore had to be removed from the study as no injury information was obtained for these participants following the first time point. To prepare the data for the BN, missing values in the dataset were first imputed. Out of the 650 total measurements across all time points in the current study, there were 31 (4.64%) missing muscle stiffness measurements and 70 (10.48%) missing heart rate recordings. The missing data were due to technical faults in the data collection equipment and were considered to be missing completely at random. A missing rate of 15-20% has been reported to be common in psychological studies, and several techniques are available to handle missing values [@Enders2003; @Lang2014]. In the current study, the *caret* package [@Kuhn2008] was used to impute the missing values. A bagged tree model using all of the non-missing data was first generated and then used to predict each missing value in the dataset. The bagged tree method is a reliable and accurate method for imputing missing values in data and is superior to other commonly used methods such a median imputation [@Kuhn2008].

A median split technique was used to discretise the data used in the network into “Low” and “High” levels. All variables apart from negative and total life events were approximately normally distributed (based on visual inspection, see supplementary material Sx) and required no further transformation prior to the median split. For the LESCA questionnaire data, a cumulative total of the current, and previous time points was calculated at each time point to account for the potential continuing effect of the life events experienced by athletes over time. Given the limited support for a relationship between positive life events and injury [@Williams2007], only negative and total life events were included in the network. Cumulative negative, and cumulative total life event scores at each time point were first log scaled so distributions were approximately normal, and then binarised using the median at each time point (nlelg and tlelg respectively). In addition to the log scaled cumulative values, an untransformed NLE score from the first time point was included as an additional variable based on previous literature that indicates this variable should have a strong relationship with injury outcome [@Ivarsson2017].

#### Network structure

To obtain the network structure, several steps were taken to ensure that both a theoretically realistic network, and a network that was an appropriate fit to the collected data, was used for inference. Prior knowledge about the network structure was included by providing a list of arcs that are always *restricted* from being in the network (blacklist), and a list of arcs that are always *included* in the network (whitelist). Additionally, there are several scoring functions such as Bayesian Information Criteria (BIC) and Bayesian Dirichlet equivalent uniform (BDeu) that can be used to compare network structures with certain nodes and arcs included or excluded [@Scutari2014]. To account for the repeated measures design and to maximise the use of the data, pairs of complete cases (e.g., participants who completed T1 + T2, and T2 + T3) were used in a two-time Bayesian network (2TBN) structure [@Murphy2002]. In the 2TBN, variables measured T2 could depend on variables measured at T1 (e.g., T1 T2) and variables measured at T3 could depend on variables measured at T2 (e.g., T2 T3). However, arcs were blacklisted between T2 T1 and T3 T2 to preserve the order in which data was collected. Variables were separated into two groups; “explanatory”, for variables that did not change during the study (e.g., gender), or “independent”, for variables that were measured at each time point and could vary during the study. Independent variable names were suffixed with \_1 for time point T, and \_2 for time point T+1 (e.g., T1\_1 T2\_2 and T2\_1 T3\_2). Formatting the data in this way meant participants who completed T1 and T2, but did not complete T3, could still be included in the analysis. Table provides an example of the formatted data. Participants 1 and 3 have complete data, and therefore have two rows of data each representing variables from T1 T2 and T2 T3, respectively. Participant 2 did not complete the final data collection at T3 and therefore only has one row of data representing the variables collected at T1 and T2. In addition to the blacklisted arcs between T2 T1 and T3 T2, the direction of arcs was restricted between independent variables and explanatory variables (e.g., independent explanatory); however, arcs were not restricted between explanatory independent variables. Finally, arc direction was restricted between specific nodes within the explanatory variables. Arcs from clevel gender, nlebase gender and nlebase ind\_team were included in the blacklist, as arcs in these directions did not make logical sense. All subsequent models used the same blacklist.

#### Preliminary network structures

Prior to the final network structure presented in the results section, several structures with different combinations of variables were explred. Networks were learned using a Tabu search algorithm [@Norvig2009] and BIC was used to compare different models. A higher BIC value indicates the structure of a DAG is a better fit to the observed data [@Scutari2014]. BIC values for each combination of variables of interest are reported as the combination of variables with the highest BIC value, followed by the relative scores of the other variables in the model.

Initially, both negative life events and total life events were included in the network structure, however, the network score was improved when only nlelg or tlelg was included (highest BIC value = nleleg, BIC values relative to nlelg; tlelg only = -83.64, tlelg and nlelg = -218.31). Additionally, despite strong evidence in the literature that both negative and total life event stress are related to injury occurrence [@Williams2007], network structures learned using the Tabu search algorithm failed to identify a relationship between NLE and injury or TLE and injury in the data. Given that nlelg provided the highest network score, and there is a stronger relationship between negative life events and injury in the literature, an arc was whitelisted between nlelg\_1 and injured\_1 and nlelg\_2 and injured\_2 in the final network structure. Total life event score was not included in the final structure.

The subscales representing the BAS (RR, RI, GDP and I) showed limited connection to other variables in the network. Therefore, several models were run with each scale individually to find the scale that resulted in the highest BIC value (values are shown relative to the highest value). RI provided the highest BIC value, compared to RR (-9.07), GDP (-15.06) and I (-16.33). Including all the variables (RR, RI, GDP and I) resulted in a significantly lower score -894.33) indicating that including all the variables was not beneficial to the model structure and did not offset the cost of the additional parameters. Therefore, only RI was included in the final structure. Finally, both total score and asymmetry were included in the initial network. However, visual inspection of the network revealed no arcs between bal\_asym\_1 or bal\_asym\_2 and any other node in the network. Therefore, balance asymmetry was removed from the final network structure. To summarise, Table includes the variables that were included in the final network structure.

Preliminary network structures also revealed strong dependencies between the same variables at subsequent time points. For example, the probability that stiffness\_1 and stiffness\_2 were both “High”, or both “Low” was approximately 80%. Including the arcs between the same variables from X\_1 X\_2 did not provide any theoretically meaningful information to the network structure as the majority of participants would be either be in a “Low” or “High” state for each pair of variables in the network. Therefore, these arcs were blacked listed from the network.

To obtain the final network, the appropriate blacklist and whitelists were provided and a Tabu search algorithm identified the remaining structure of the network. The final network structure was obtained by averaging 1000 bootstrapped models [@Efron1993] to reduce the impact of locally optimal, but globally suboptimal network learning, and to obtain a more robust model [@Olmedilla2018]. Arcs that were present in at least 30% of the models were included in the averaged model. The strength of each arc was determined by the percentage of models that the arc was included in, independent of the arc’s direction. An arc strength of 1 indicates that the arc is always present in the network, with the value decreasing as arcs are found in fewer networks. In the respective study arcs above 0.5 were considered “significant” with arcs below 0.5 and above 0.3 “non-significant” [@Scutari2013]. Arcs below 0.3 were not included in the model. The full table of arc strengths for the first and second network are available in S2 Table and S3 Table respectively.

#### Network Inference

Conditional probability queries (CPQ) were used to perform inference on both network structures. To conduct a CPQ, the joint probability distribution of the nodes was modified to include a new piece of evidence. The query allows the odds of a particular node state (e.g., injured\_1 = “injured”) to be calculated based on the new evidence. CPQs were performed using a likelihood weighting approach; a form of importance sampling where random observations are generated from the probability distribution in such a way that all observations match the evidence given in the query. The algorithm then re-weights each observation based on the evidence when computing the conditional probability for the query [@Scutari2014]. Inference was first performed on arcs that had a strength greater than 0.50 between the explanatory variables and independent variables and between different independent variables in the network. Of particular interest in the current study were the variables that were connected to “injured” nodes. To examine the variables that were associated with injured nodes in the network, the Markov blanket of “injured\_1” and “injured\_2” were examined. A Markov blanket contains all the nodes that make the node of interest conditionally independent from the rest of the network [@Fuster-Parra2017]. CPQ’s were used to determine what effect the variables in the Markov blanket of injured nodes had on the probability of the injured node being in the “injured” state.

### Second network

#### Data preparation

For the second network, continuous variables suffixed with \_1 were subtracted from variables suffixed with \_2 to obtain the difference between variables measured at time points T1 T2 and T2 T3. Independent variables were then standardised (mean-centred at 0) to allow relative changes between variables to be compared. The “injured” variable was also modified to represent whether a participant had sustained an injury at any point over the duration of the study or were healthy for the duration of the study. Setting the data up in this way enabled the construction of a network that explicitly modelled the *amount* of change within variables between time points, as opposed to the first network that would only have captured changes when the median threshold was crossed from “Low” to “High”.

#### Network structure

Similar to the first network, blacklists were used to prevent arcs from independent variables explanatory variables. In addition, “nlebase” variable was dropped from the list of explanatory variables to allow the *changes* in negative life events to be the only life event variable in the network. The final network was obtained using the same approach as the first network.

#### Network Inference

Conditional probability queries (CPQ) were again used to perform inference the network structures. The Markov blanket of the “injured” node was of particular interest, and the probability of injury was investigated with combinations of variables in the Markov blanket below the mean change, at the mean change and above the mean change. Initial visual inspection of the network structure also revealed arcs between BIS FFFS and HRV FFFS. To investigate this relationship further, random samples were generated for BIS, HRV and FFFS based on the conditional distribution of the nodes included as evidence in the query. The samples were then examined using a Bayesian linear regression models with the *brms* package [@Burkner2017a] to determine the relationship between these nodes. Weakly regularising priors (normal prior with mean of 0 and standard deviation of 5) were used for all parameters in the model.

1. The term “query” in relation to Bayesian Networks stems from Pearl’s expert systems theory (1988). A query can be submitted to an expert (in this case, the network is the expert) to get an opinion, the expert then updates the querier’s beliefs accordingly. Widely used texts on Bayesian Network analysis [@Koller2009] have widely adopted the terminology in favour of that used in traditional statistics. [↑](#footnote-ref-1)