

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/330293106>

# Developing Athlete Monitoring Systems in Team Sports: Data Analysis and Visualization

Article in *International Journal of Sports Physiology and Performance* · January 2019

DOI: 10.1123/ijspp.2018-0169

CITATIONS

30

READS

13,539

4 authors:



**Heidi R Thornton**

La Trobe University

56 PUBLICATIONS 999 CITATIONS

[SEE PROFILE](#)



**Jace A. Delaney**

Victoria University Melbourne

40 PUBLICATIONS 863 CITATIONS

[SEE PROFILE](#)



**Grant M Duthie**

Australian Catholic University Strathfield NSW

83 PUBLICATIONS 3,177 CITATIONS

[SEE PROFILE](#)



**Ben Dascombe**

La Trobe University

163 PUBLICATIONS 4,211 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Peak demands [View project](#)



Collaborative development of perceptual-cognitive tests for research and teaching [View project](#)

*Note.* This article will be published in a forthcoming issue of the *International Journal of Sports Physiology and Performance*. The article appears here in its accepted, peer-reviewed form, as it was provided by the submitting author. It has not been copyedited, proofread, or formatted by the publisher.

**Section:** Brief Review

**Article Title:** Developing Athlete Monitoring Systems in Team-Sports: Data Analysis and Visualization

**Authors:** Heidi R. Thornton,<sup>a,b</sup> Jace A. Delaney,<sup>c</sup> Grant M. Duthie,<sup>d</sup> and Ben J. Dascombe<sup>e</sup>

**Affiliations:** <sup>a</sup> La Trobe Sport and Exercise Medicine Research Centre, La Trobe University Bundoora Campus, Melbourne, Victoria, 3086, Australia; <sup>b</sup> Newcastle Knights Rugby League Club, Mayfield, New South Wales, 2304, Australia; <sup>c</sup> Institute of Sport Exercise and Active Living, Victoria University, Footscray Campus, Melbourne, Victoria, 8001, Australia; <sup>d</sup> School of Exercise Science, Australian Catholic University, Strathfield, New South Wales, 2137, Australia; <sup>e</sup> Applied Sports Science and Exercise Testing Laboratory, University of Newcastle, Ourimbah, NSW, 2258, Australia.

**Journal:** *International Journal of Sports Physiology and Performance*

**Acceptance Date:** December 9, 2018

©2019 Human Kinetics, Inc.

**DOI:** <https://doi.org/10.1123/ijsp.2018-0169>

**Title:** Developing athlete monitoring systems in team-sports: data analysis and visualization

**Authors:** Heidi R. Thornton,<sup>a,b\*</sup> Jace A. Delaney,<sup>c</sup> Grant M. Duthie,<sup>d</sup> & Ben J. Dascombe<sup>e</sup>

**Brief running head:** Athlete monitoring systems in team-sports

**Submission Type:** Invited review

**Laboratory:** La Trobe Sport and Exercise Medicine Research Centre, La Trobe University

**Institutions:**

<sup>a</sup> La Trobe Sport and Exercise Medicine Research Centre, La Trobe University Bundoora Campus, Melbourne, Victoria, 3086, Australia;

<sup>b</sup> Newcastle Knights Rugby League Club, Mayfield, New South Wales, 2304, Australia;

<sup>c</sup> Institute of Sport Exercise and Active Living, Victoria University, Footscray Campus, Melbourne, Victoria, 8001, Australia;

<sup>d</sup> School of Exercise Science, Australian Catholic University, Strathfield, New South Wales, 2137, Australia

<sup>e</sup> Applied Sports Science and Exercise Testing Laboratory, University of Newcastle, Ourimbah, NSW, 2258, Australia

**Corresponding Author:**

Miss Heidi Rose Thornton  
La Trobe Sport and Exercise Medicine Research Centre,  
La Trobe University, Victoria, 3086  
Email: [19088233@students.latrobe.edu.au](mailto:19088233@students.latrobe.edu.au)  
Phone: +61 423 067 434

**Abstract word count:** 194

**Text-only word count:** 4836

**Number of tables:** 1

**Number of figures:** 2

## ABSTRACT

**Purpose:** Within professional team sports, the collection and analysis of athlete monitoring data is common practice in the aim of assessing fatigue and subsequent adaptation responses, examining performance potential as well as minimizing the risk of injury and/or illness. Athlete monitoring systems should be underpinned by appropriate data analysis and interpretation, in order to enable the rapid reporting of simplistic and scientifically valid feedback. Using the correct scientific and statistical approaches can improve the confidence of decisions made from athlete monitoring data. However, little research has discussed and proposed an outline of the process involved in the planning, development, analysis and interpretation of athlete monitoring systems. This review discusses a range of methods often employed to analyze athlete monitoring data to facilitate and inform decision making processes. **Conclusions:** There is a wide range of analytical methods and tools that practitioners may employ within athlete monitoring systems, as well as several factors that should be considered when collecting these data, methods of determining meaningful changes and various data visualization approaches. Underpinning a successful athlete monitoring system is the ability of practitioners to communicate and present important information to coaches, ultimately resulting in enhanced athletic performance.

**Key words:** Fatigue, individual responses, meaningful change, communicate, reporting

## INTRODUCTION

Athlete monitoring data provides useful information as to whether athletes are responding appropriately to imposed training and competition demands.<sup>1</sup> Evaluating training monitoring data is crucial to ensure that athletes are exposed to sufficient training to prepare them for the requirements of competition, whilst ensuring the athletes are appropriately adapting to the training program. Together, this process should assist in minimizing the risk of undergoing excessive workloads,<sup>1</sup> and maximize the performance potential of athletes. In team sports, sport scientists and support staff often collect a wide range of athlete monitoring data that relate to external and internal workloads that their athletes have undergone, as well as physiological and psychological assessments of fatigue.<sup>2</sup> When appropriately collected and interpreted, these data should inform decision-making processes regarding the planning and manipulation of training. However, limited research has explored how key variables should be selected, how such data should be analyzed, how meaningful information can be extracted, and what methods present these data most effectively.

Applied sport scientists are faced with an extensive range of monitoring technologies, therefore selecting and utilizing the most appropriate variables that are sensitive to fatigue or answer the questions of coaches is necessary. As previously explained,<sup>1</sup> training monitoring data should be interpreted with knowledge of the tool's validity, reliability as well as any implications for injury, illness and expected player performance. Other factors such as measurement error, individualization and discretization of data should be considered, and as such are discussed. In this review, a range of data analytical methods which are commonly used in applied and research settings have been presented, which can be used to determine and interpret meaningful changes in data, assisting in informing decision making. Lastly, various data visualization concepts and methods have been discussed, which aim to enhance the success of the monitoring system through effective communication of relevant information.

The overall purpose of this review is to provide a methodological outline that may assist team sport practitioners to develop and employ simple and effective feedback for coaches and athletes. An overview of the framework for developing an athlete monitoring system is depicted in Figure 1, with each step detailed within the review.

### **Step 1. Considerations for athlete monitoring in team sports**

Collecting and understanding the useful and informative variables is imperative to a successful athlete monitoring system.<sup>3</sup> The factors discussed below should be considered prior to the collection of any data, as per the process demonstrated in Figure 1. The ability to collect and access athlete data has rapidly expanded with the increased availability of monitoring technologies, for example, data from Global Positioning System (GPS) devices, accelerometers/gyroscopes/magnetometers, heart rate monitors, force plates, linear force transducers and self-report measures. Given the large quantity of data, practitioners are required to select which data best help answer the questions of coaches and athletes.<sup>3</sup> Practitioners should consider several factors prior to collecting athlete monitoring data, such as how these data will be collected (e.g. logistics, resources), what the potential limitations are of certain variables (e.g. validity, reliability),<sup>4</sup> as well as how these data can be effectively reported back to coaches and athletes. These variables should be appropriate in assessing the outcomes of the training program (e.g. performance), but also be useful in the prescription of training.<sup>3</sup> To identify relevant variables that are suitable, practitioners may wish to adopt the “working fast-working slow” concept.<sup>5</sup> Using this concept, applied research could identify variables that are associated with the programs outcome measures,<sup>6</sup> for assessing individual changes in fitness, and injury/illness rates.<sup>7</sup>

The importance of individualized monitoring is apparent, given the varying responses to a given training stimulus, potentially influencing the rate of adaptation.<sup>8</sup> Inter-athlete

variability to training may result from various factors, including age, gender, training history, fitness level, psychological status and genetics.<sup>9</sup> Importantly, these factors affect how athlete monitoring data is analyzed, interpreted and presented. When depicting changes in an individual's data (e.g. fitness testing), practitioners should present within-athlete changes. Furthermore, when interpreting changes in such responses, individualized methods may be used, such as Z-scores to highlight the relative change in the variable. Within research the same importance of individualization is applicable.<sup>10</sup> As individual characteristics may modify the effect of the treatment, researchers should acknowledge and attempt to measure varying individual characteristics.<sup>10</sup> Although often this is not possible, as controlled trials require large sample sizes, or repeated measures to be averaged to compensate for large measurement errors.<sup>10</sup> Means and standard deviations (SD) should be reported, as these depict the variation in the responses, reflecting the effect of the treatment between individuals. The SD at baseline may be used to assess the magnitudes of effects and individual responses by standardization.<sup>10</sup> These data may be reported at the individual or group level, where the mean and SD of the change scores for each experimental group is depicted.<sup>10</sup> Additionally, confidence intervals (CI) may be included, as CI are effective in providing the true value of the variance of individual responses,<sup>10,11</sup> although the specific use of CI will be discussed later in this review.

Separately, an example of accounting for individual characteristics of external training loads through GPS analysis, is the use of individualized speed thresholds. Individualized thresholds consider athletes' speed capacities (rather than absolute thresholds, e.g.  $5 \text{ m}\cdot\text{s}^{-1}$ ) when determining high-speed running distance,<sup>12</sup> as varying capacities result in varying speeds at which high-intensity running begins.<sup>12</sup> Proprietary GPS software typically allows for the customizability of these thresholds, that may be determined using measures of anaerobic capacity,<sup>13</sup> intermittent exercise capacity,<sup>14</sup> maximal aerobic speed,<sup>15</sup> maximal running speed, or a combination of measures.<sup>15,16</sup> Whilst individualized thresholds have merit when planning

programs and when quantifying training load, there limitations to consider when deriving these thresholds.<sup>16</sup> As the determination of individualized thresholds requires testing, practitioners must conduct this frequently within the training program. If this isn't possible, seasonal variations in physical fitness will not be assessed, therefore the threshold will not be reflective of the athletes' true capacity. Therefore, absolute thresholds may be more suitable in some contexts, perhaps in sports where there isn't a large variation in physical capacities of athletes, but also permits the longitudinal assessment of the physical demands of competition. Within research, absolute thresholds permit comparisons between sports and/or competitions,<sup>16</sup> however variations in GPS manufacturers, sampling frequencies and data processing methods influence such comparisons.<sup>16,17</sup> Taken together, practitioners should consider which method is most appropriate to evaluate data that informs training prescription.<sup>16</sup>

Categorizing continuous data into discrete categories (discretization) is common in applied settings, for example using speed and acceleration/deceleration-based thresholds and Z-score categories.<sup>18</sup> The discretization of data may assist in summarizing and presenting data and is therefore prevalent in sport, however this practice is receiving attention as an issue.<sup>18-20</sup> The discretization of data results in the loss of statistical power when there are numerous 'categories', as the relation between the predictor and the outcome is reduced.<sup>19</sup> Specific to injury risk, discretization assumes that each individual has the same risk within each category, and the within-category variation is ignored.<sup>18</sup> Evidence has shown that discrete models were inferior to continuous models for fitting the injury risk profiles of Australian Football and soccer athletes.<sup>18</sup> Given these factors, practitioners may wish to keep data as continuous until more appropriate methods are available, although their use is still common (e.g. speed thresholds, flagging systems). Alternative methods such as spline regression or fractional polynomials as suggested in research may be used, although are beyond the scope of this review.<sup>18</sup> In conclusion, practitioners must consider a range of factors prior to collecting athlete



monitoring data, and make informed decisions depending on the context of the sport and the aims of the monitoring system.

## **Step 2. Methods of analyzing athlete monitoring data**

The methods employed to analyze athlete monitoring data may depend on factors such as the analytical skills of the sport scientist, resources available as well as the philosophies of the coach and other staff. There are many tools available for practitioners regardless of skills or resources, which may allow meaningful information to be derived from the data, allowing simple and effective feedback to be produced. Firstly, practitioners should consider where monitoring data will be stored and accessed by relevant stakeholders, as this may influence the process of how data will be further analyzed. For example, many professional sporting teams utilize commercial athlete monitoring software (AMS), which can be a secure and effective method to store, analyze and present data. Commercial AMS can vary greatly in cost due to various additional features which may assist in the analysis of data. Practitioners may prefer to conduct their own analyses using programs such as Microsoft Excel<sup>®</sup> (Microsoft, Redmond, Washington) or R Studio software.<sup>21</sup> This may be due to financial reasons, limited buy-in of other staff when using commercial AMS, or the skills necessary to develop and manage the AMS. Despite these factors, this review has presented some methods of analyzing monitoring data, irrespective of where these may be conducted.

Specific to training load data, there are numerous methods of calculating cumulative loads for pre-defined periods, such as 7, 14 and 21-days. Compared to pre-defined weekly periods (e.g. Monday to Sunday), rolling totals may be a useful method, assisting in overcoming the influence of varied match recovery periods between games that is common in team sports.<sup>22</sup> Using rolling loads, the acute to chronic workload ratio (ACWR) has been developed and investigated in a number of team-sports.<sup>23-25</sup> This workload index provides an

indication of where the athlete’s acute (e.g. 7 days) workload is relative to chronic period (e.g. 28 days), providing a ratio score.<sup>25,26</sup> Despite the prevalence of these specific time frames or ‘windows’ in research, the most appropriate windows should be reflective of the sport, the structure of competition and training requirements,<sup>27</sup> but also the association with injury and/or performance. For example, in the National Basketball Association competition, given the high frequency of games (> 1 game per week), a shorter ‘acute’ period (e.g. 3 days) may be most appropriate. Practitioners may wish to investigate which windows are most applicable within their environment and be aware that these windows may vary depending on the variable being measured.<sup>27</sup> To determine which variables and windows are most applicable, the methods employed within applied research<sup>27</sup> are recommended.

The ACWR has been utilized to examine injury risk in a variety of team-sports.<sup>24-26</sup> For example, in Australian Rules Football,<sup>26</sup> during the in-season period, an ACWR of >2.0 for high-speed distance was associated with a 6 to 12 times greater risk of injury than an ACWR of <0.49 (relative risk =11.62). Whilst the ACWR is a simplistic tool to identify spikes in workloads and potential injury risk,<sup>25</sup> there are a number of considerations which have been highlighted<sup>28,29</sup> and alternative methods provided.<sup>30</sup> Specifically, the use of rolling loads and the ACWR fail to account for the decaying nature of fitness and fatigue effects over time. To account for this effect, the exponentially-weighted moving averages (EWMA) have been proposed.<sup>30</sup> Specifically, this method involves assigning a decreasing weighting factor for each older load value, where the EWMA for a given day is given by:

$$EWMA_{today} = Load_{today} * \lambda_a + ((1 - \lambda_a) * EWMA_{yesterday}) \quad (1)$$

Where  $\lambda_a$  is a value between 0 and 1 that represents the degree of decay, with higher values discounting older observations at a faster rate. The  $\lambda_a$  is given by:

$$\lambda_a = 2/(N + 1) \quad (2)$$

Where  $N$  (the chosen time decay constant) is typically 7 days for acute and 28 days for chronic time periods.<sup>30</sup> As per the rationale explained previously, the choice of time period (e.g. 3 or 7-days) should be determined by the practitioner. The EWMA approach may be better suited to team-sports such as rugby league,<sup>31,32</sup> and Australian Rules Football<sup>33</sup> for the modelling of training loads, as a greater weighting to the most recent workload undergone, therefore accounting for the decaying nature of fitness.<sup>33</sup> The use of EWMA loads as opposed to rolling loads can then be used to calculate the ACWR,<sup>26</sup> but also may be used to simply describe the training load undergone over a set period of time. Despite the proposed benefits of using both the ACWR and EWMA loads, there are limitations and considerations when using these methods.

Potential limitations with the ACWR may include the locomotor profiles of the athletes, varying fitness of athletes, and the inability to collect consistent data across a season.<sup>29</sup> Specifically, in regards to high-speed running thresholds, as per the limitations associated with individualizing thresholds,<sup>16</sup> training may be adjusted based on an athlete’s high-speed running ACWR being too high (depending on the ACWR that the practitioner deems as being an ‘injury risk’). However, the athlete’s fitness may have improved substantially since the last testing period causing their threshold to be inaccurate, therefore potentially predisposing the athlete to injury risk if inadequate training is undertaken. Additionally, athletes being intermittently absent from training due to international competition, access to training load data may be unavailable or inconsistent with previously collected data, potentially affecting the use of the

ACWR.<sup>29</sup> Further, issues pertaining to the calculation of workload ratios have been highlighted.<sup>28</sup> Given that one aspect of the ACWR calculation is that the acute load also constitutes a substantial part of the chronic load, this results in “mathematical coupling” between two variables.<sup>28</sup> This mathematical coupling raises the possibility that that inferences made from such data may be compromised by spurious correlations.<sup>28,34</sup> Spurious correlations are relationships that occur between two variables, irrespective of biological and physiological association.<sup>28,34</sup> Recently, research has proposed the “uncoupled” ACWR, which is defined as the ratio of the most recent week of work with the average of the three preceding weeks.<sup>35</sup> This research<sup>35</sup> presents a formula to convert the traditional “coupled” ACWR to the newly proposed “uncoupled” ACWR, although to date there is limited evidence available of its use to assess injury risk. Given the range of methods available to calculate workload progression, practitioners should select the tool that is deemed as most relevant in the context of identifying changes in athlete status.

An alternative method of determining changes in athlete monitoring data is using standardized scores, or Z-scores. Z-scores are prevalent in analyzing daily monitoring tools such as subjective wellness, where informed decisions need to be made quickly and effectively. To calculate a Z-score, if  $x$  is an observation from a distribution that has a mean ( $\mu$ ) and SD ( $\sigma$ ), the Z-score of  $x$  is;

$$Z = (x - \mu) / \sigma \quad (3)$$

This Z-score value represents how many SD the original observation is from the mean, and in which direction this falls.<sup>36</sup> An assumption of Z-scores is that the original data is normally distributed,<sup>37</sup> allowing accurate interpretation of the observation, therefore practitioners should test for normality. Although this is beyond the scope of this review, methods of determining data distribution are explained in research.<sup>37</sup> As the training and

competition requirements vary across a season, (e.g. during pre-season training may be maximized) athletes’ responses (e.g. muscle soreness) will vary. Therefore, when calculating Z-scores using historical data, a select time period of which data is included in the Z-score calculation may be useful. For example, during early pre-season, a 3-week block of historical data may be used to calculate a Z-score, as the athletes are still adapting to training (e.g. high perceived soreness). Alternatively, practitioners may select specific blocks (e.g. 3-weeks whilst on tour overseas), or 6-week rolling averages to account for changes in fitness across the season. Depending on the aims of the monitoring program or the desired sensitivity of their measures, practitioners may wish to alter what corresponds as a ‘red flag’ based on these Z-scores. For example, Table 1 demonstrates the use of Z-scores on subjective wellness data, where a Z-score of -1.5 to -1.99 (SD) is used as a ‘yellow flag’, and a Z-score of  $< -2$  constitutes a ‘red flag’. Often it can be difficult for coaches to comprehend the Z-score system, therefore standard tens (STEN) scores are an effective tool (anecdotally) for expressing Z-scores on a 1 to 10 scale. To convert a Z-score to a STEN score, the formula below can be used; are calculated as;

$$\text{STEN} = (2 Z) + 5.5 \quad (4)$$

Specifically, each STEN unit corresponds to a SD of 0.5, whereby a Z-score of 0 (e.g. no change from mean) is represented by a STEN score of 5.5. A STEN score of 1 reflects a Z-score of  $\leq -2$ , and a STEN score of 10 corresponds to a Z-score of  $\geq 2$ . For example, a Z-score of -1.8 (poor or below mean) corresponds to a STEN score of 1.9, where the practitioner may decide on specific ‘flags’ as per that of Z-scores. An example of STEN scores being used to represent seven-day EWMA loads for one athlete is show in Figure 1, where in practice various colors could be used on a figure to represent STEN scores (e.g. 1-3; very low, 3-5; low, 5-8; optimal; 8-9, high; 9-10, very high). By using standardized scores, missing data (e.g.

representative duties, injured or low compliance) doesn't affect group means. Another benefit of using standardized scores is that practitioners may present various monitoring variables using the same scale, which has substantial benefit in interpreting and communicating findings.

### **Step 3. Determining meaningful changes in data**

To make informed decisions from athlete monitoring data, a meaningful change in responses (beyond normal or random variability) must be established.<sup>38</sup> There are multiple ways to determine what constitutes a meaningful change in athlete monitoring data, including SD, typical error, effect sizes, smallest worthwhile change (SWC), coefficient of variation (CV) and magnitude-based inferences (MBI).<sup>38-40</sup> These statistical processes represent alternatives to null-hypothesis significance testing (NHST), are well-established analytical approaches in applied settings, and can be presented in a practical manner so that they are understood by coaches and players.<sup>3</sup> Importantly, practitioners should be aware that these data and the meaningful information derived should be used simply to inform coaches regarding athlete status, and should not dictate decision-making process.

Standard deviations are a useful method of demonstrating the spread of data, simply calculated as the root mean square of the differences from the mean value. Further, SD can often be expressed as a percent of the mean, also known as the CV which presents an important measure of reliability.<sup>41</sup> It is important to consider the typical error of measurement when interpreting athlete data, as this error represents the random variation that may occur due to biological and technological error.<sup>42</sup> Assessing the typical error of measurement is commonly determined from test-retest reliability analysis. It is quantified as the SD of the difference between the athletes' observed values, divided by  $\sqrt{2}$ . Comparing the magnitude of the change in data to the typical error may reveal whether a “real” change has occurred.<sup>42</sup> For example, when determining a change in a fitness test score such as a 40m sprint, the typical error may

be 0.09 seconds. Therefore, any change in 40m sprint time of  $<0.09$  seconds would not constitute a ‘real’ change in performance, whereas changes  $>0.09$  seconds would represent an actual change. This method is useful when using tests that are used rarely, where more ‘noise’ or variation in the data will occur. Whilst the typical error is simple and effective in determining whether a true change in fitness score has occurred outside of the variability in testing results, there is no evidence of whether the change is “worthwhile” in the context of performance.<sup>42</sup>

The SWC that is also referred to as the smallest clinically important difference can be defined as a threshold, in which if a change in a given variable exceeds, is interpreted as meaningful.<sup>3,39</sup> In team-sports, the SWC presents as an effective method to determine meaningful changes in data (e.g. improving fitness). The SWC can be calculated as;  $0.2 \times$  the between-subject SD,<sup>39</sup> that is based on the concepts of Cohen’s  $d$  effect size.<sup>42</sup> For a variable to be considered suitable in detecting the SWC using this method, the typical error of measurement should be lower than the SWC.<sup>43</sup> This ensures that practitioners can be confident in detecting changes in responses over a period of time rather than identifying typical variation in the test.

When comparing changes in means from independent groups, standardized differences can be computed to determine the magnitude of difference.<sup>44</sup> A common method of determining the effect between the two group means is using Cohen’s  $d$ .<sup>45</sup> Using this method, the difference between the group means is expressed as a factor of the between-athlete SD, that can be described using an alternate interpretation.<sup>39</sup> In this scale, the magnitude of effects can be described as;  $<0.2$  trivial,  $0.21 - 0.60$  small,  $0.61 - 1.20$  moderate,  $1.21 - 2.0$  large and  $>2.0$  very large.<sup>39</sup> This scale is more robust than that of Cohen,<sup>45</sup> as there are additional levels or thresholds, and has commonly been utilized in team-sport research and applied practice.<sup>3</sup> However, it is important that effect sizes be accompanied by CI to provide a true range of the effect statistic.<sup>11</sup> The CI statistic defines the range representing the uncertainty in the true value

of an effect statistic given the sample size, and the range of values are likely to include the population's true value if replicated.<sup>11,43</sup> To derive CI, a customized spreadsheet is available,<sup>46</sup> where CI are derived from the  $p$  value, the degrees of freedom, and the effect statistic of interest. Typically, CI are set at 90%, meaning that the chances are 90% that the CI encloses the true effect statistic.<sup>46</sup> For example, a correlation between two internal load measures may be  $r = 0.68$ , and using 90% CI, the true value may lay between 0.60 and 0.76. When used in conjunction with an effect statistic, range of the CI will indicate the magnitude of uncertainty of the statistic.<sup>11</sup> This approach is useful in making MBI, identifying where the CI lies in relation to threshold values for substantial effects, although this will be explained in further detail.<sup>11,39</sup>

MBI are an alternative statistical approach to NHST that has largely been driven by Hopkins and colleagues.<sup>11,39</sup> This approach uses two simple concepts to present inferences using CI. Firstly, changes or differences in data are compared to a threshold representing the SWC, where the effect can then be determined as being “harmful”, “trivial” and “beneficial” depending on where the effect statistic and CI lie.<sup>11</sup> Unclear effects can occur, where the uncertainty of the effect crosses both positive and negative thresholds. The probability of the effects can be reported quantitatively, where percentage probability of the effect being lower, similar or higher than the SWC are reported (e.g. 20/10/70). These percentage chances are associated with qualitative descriptors; 0-0.5%, most unlikely, 0.5-5%, very unlikely; 5-25%, unlikely; 25-75%, possibly; 75-95%, likely; 95-99.5%, very likely and >99.5%, most likely.<sup>39</sup> MBI are easily calculated using a specially designed spreadsheet freely available online,<sup>46</sup> and are useful in the practical interpretation of findings,<sup>39</sup> and in improving data visualization.<sup>3</sup> As highlighted,<sup>3</sup> MBI can be incorporated into graphical reports by shading trivial areas, displaying CI, and including symbols to depict substantial changes where necessary.



Despite the advocacy towards the use of MBI in sports science,<sup>40</sup> several statisticians have criticized MBI, relating to a lack of theoretical framework and that it creates unacceptably high false positive rates (Type 1 error).<sup>47,48</sup> The recent opposing views on MBI may have been prompted by the conclusions of research,<sup>49</sup> where MBI was represented as superior to the traditional NHST approach in terms of inferential error rates, rates of publishable outcomes with suboptimal sample sizes, and publication bias.<sup>49</sup> In response, Sainani<sup>48</sup> proposed that NHST is superior to MBI, stating that MBI “exhibits undesirable systematic behavior and should not be used”. It was concluded that MBI should be replaced with a fully Bayesian approach, or a one-sided NHST alongside the corresponding CI would achieve the objectives of MBI, whilst correctly controlling Type 1 error.<sup>48</sup> An alternative method has been presented,<sup>50</sup> demonstrating a fully Bayesian approach that can be used with small sample sizes whilst providing directly interpretable outcomes, more specific to applied research. Whilst it is beyond the scope of this review to provide extensive detail on this alternative approach,<sup>50</sup> other research provides further information that may also assist practitioners in selecting which method is most appropriate in a range of environments.<sup>43</sup>

It is important to note that identifying meaningful changes in athlete data is only a small step in the athlete monitoring process. It is the actions of practitioners following these findings in the data that results in a successful monitoring system. For example, practitioners may raise issues regarding substantial changes in data, however, unless this issue is actioned upon (e.g. training be modified), the efficacy of the monitoring system is limited.

#### **Step 4. Data visualization and communication**

As explained previously,<sup>3</sup> the ability of practitioners to deliver the information obtained by the monitoring system is crucial in the overall success of the program. It is this skill of communicating information and by gaining the trust of coaches, that as practitioners, our data

can influence decisions being made. The ability to deliver data to athletes and coaching staff is optimized through visually appealing means that are simple and informative. Depending on the data, the method of presentation may differ, with various types of graphs (e.g. line graphs, bar charts, pie graphs) or tables to present data. When developing these resources particularly for regular distribution (daily/weekly), there are various considerations that may ensure the simple and informative relay of data.<sup>3</sup> For example, selecting variables that inform decision-making (e.g. distance, high-speed distance), ensuring that unnecessary noise is removed (such as decimal places),<sup>51</sup> as well as making sure that text is readable, and formatting allows for easy interpretation. Presenting changes in athletes’ data (e.g. pre/post testing results) can be accomplished using raw descriptive data but extend to include further levels of data analysis such as MBI, group and within-subject changes in data. Post-analysis conditional formatting using Microsoft Excel® (Microsoft, Redmond, Washington) is a simple method of highlighting important information, and may be used to effectively implement ‘traffic light’ systems that convey the athlete monitoring information in an easily interpretable manner.<sup>38</sup> Another example of this may be taken from when using STEN scores to depict training loads to coaches, as shown in Figure 2, where ‘bands’, such as very low, low, optimal, high and very-high may be used. To develop visually appealing reports and graphics, Excel® (Microsoft, Redmond, Washington) is a useful and simple tool, but additional programs such as Tableau® (Seattle, Washington), R statistical software<sup>21</sup> and Power BI® (Microsoft, Redmond, Washington) are also available, and may be useful in producing powerful, effective visualizations.

When presenting data, increasing the transparency in how data is displayed is important.<sup>52</sup> Figures should be presented that support the key findings, and allow readers to evaluate the data. Selecting the most appropriate method to visualize the data is important, as often certain graphs can cause certain characteristics of the data to be misinterpreted, such as

the distribution, paired or independent data, and outliers.<sup>52</sup> Specifically, many different data distributions can lead to a similar bar or line graph, despite the actual data suggesting different conclusions from the summary statistics.<sup>52</sup> Univariate scatterplots or dot plots are recommended, showing the raw data when there is a small sample size, or use box plots with interquartile ranges (25<sup>th</sup> and 75<sup>th</sup> percentile of the sample), where whiskers may be included to demonstrate outliers.<sup>52</sup> Violin plots are effective in demonstrating the distribution of the data in medium and large sample sizes, and bar graphs should be avoided in presenting continuous data, particularly with small sample sizes.<sup>52</sup> To create quality figures, software such as SigmaPlot (Systat Software, San Jose, California), Tableau® (Seattle, Washington), R Studio<sup>21</sup> are recommended. Meaningful changes in data may be demonstrated, including either CI or typical errors for individual data.<sup>3</sup>

### **Practical Applications**

- There are many factors practitioners should consider prior to implementing monitoring systems. These include establishing the purpose of the monitoring system, and identifying outcome variables that are valid, reliable and sensitive in examining changes in performance and illness/injury risk.
- Use appropriate analytical methods to derive meaningful information from monitoring data to assist the decision-making processes.
- Developing visual reporting tools that produce simplistic yet informative reports for coaches and athletes.
- Underpinning a successful monitoring system is the ability of the practitioner to communicate necessary information to relevant stakeholders, allowing action to be taken where necessary.

## Conclusions

There has been an exponential increase in the quantity of athlete monitoring data collected within team sports. As such, practitioners must recognize the purpose of the monitoring system and the needs of the team’s stakeholders (e.g. coaches/players) prior to collecting data. Simple, valid and reliable variables are recommended, that relate to the outcome of the program (e.g. assess changes in performance, injury/illness risk) and can be used to prescribe training. Analytical methods to analyze data may be necessary, which may be further investigated to determine if any meaningful change has occurred. The success of the monitoring system is dictated by the ability of the practitioner to communicate necessary findings to relevant staff, allowing informed decisions regarding athlete status to be made. The communication of information may be facilitated by producing attractive and informative reports that highlight the key findings of the data.

## REFERENCES

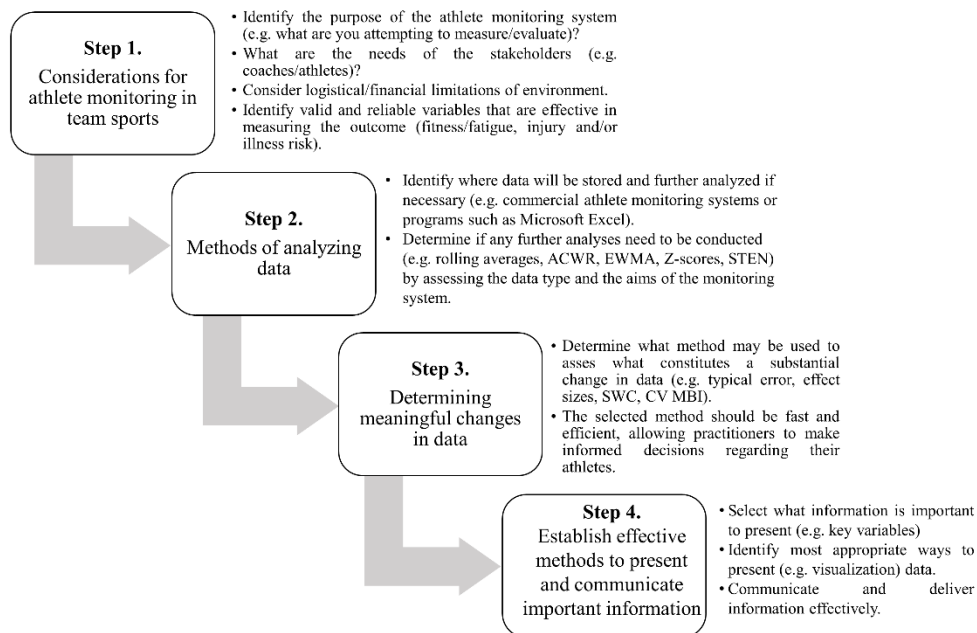
1. Thorpe R, Atkinson G, Drust B, Gregson W. Monitoring fatigue status in elite team sport athletes: Implications for practice. *Int J Sports Physiol Perform*. 2017;12(Suppl 2):S227-S234.
2. Halson S. Monitoring training load to understand fatigue in athletes. *Sports Med*. 2014;44(2):139-147.
3. Buchhiet M. Want to see my report, coach? Sport science reporting in the real world. *Aspetar Sports Medicine Journal*. 2017;6:36-42.
4. Atkinson G, Nevill A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26(4):217-238.
5. Coutts A. Working fast and working slow: The benefits of embedding research in high performance sport. *Int J Sports Physiol Perform*. 2016;11(1):1-2.
6. Thorpe R, Atkinson G, Drust B, Gregson W. Monitoring fatigue status in elite team-sport athletes: Implications for practice. *Int J Sports Physiol Perform*. 2017;12(Suppl 2):S227-S234.
7. Hopkins WG, Marshall SW, Quarrie KL, Hume PA. Risk factors and risk statistics for sports injuries. *Clinical Journal of Sport Medicine*. 2007;17(3):208-210.
8. Halson SL. Monitoring training load to understand fatigue in athletes. *Sports Med*. 2014;44 (Suppl 2):S139-147.
9. Borresen J, Lambert MI. The quantification of training load, the training response and the effect on performance. *Sports Med*. 2009;39(9):779-795.
10. Hopkins W. Individual responses made easy. *J Appl Physiol*. 2015;118(12):1444-1446.
11. Batterham A, Hopkins W. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*. 2006;1(1):50-57
12. Scott T, Thornton H, Scott M, Dascombe B, Duthie G. Differences between relative and absolute speed and metabolic thresholds in rugby league. *Int J Sports Physiol Perform*. 2018;13(3):298-304.
13. Clarke A, Anson J, Pyne D. Physiologically based GPS speed zones for evaluating running demands in Women's Rugby Sevens. *J Sports Sci*. 2015;33(11):1101-1108.
14. Buchheit M, Hammond K, Bourdarios P, et al. Relative match intensities at high altitude in highly-trained young soccer players (ISA3600). *J Sports Sci Med*. 2015;14(1):98-102.
15. Lacombe M, Piscione J, Hager JP, Bourdin M. A new approach to quantifying physical demand in rugby union. *J Sports Sci*. 2014;32(3):290-300.
16. Malone J, Lovell R, Varley M, Coutts A. Unpacking the black box: Applications and considerations for using GPS devices in sport. *Int J Sports Physiol Perform*. 2017;12(Suppl 2):S218-S226.

17. Thornton H, Nelson A, Delaney J, Duthie G. Inter-unit reliability and effect of data processing methods of global positioning systems. *Int J Sports Physiol Perform.* 2018;In Press.
18. Carey D, Crossley K, Whiteley R, et al. Modelling training loads and injuries: The dangers of discretization. *Med Sci Sports Exerc.* 2018;In Press.
19. Altman D, Royston P. The cost of dichotomising continuous variables. *Brit Med J.* 2006;332(7549):1080.
20. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol.* 2012;12(1):21.
21. *R: A language and environment for statistical computing* [computer program]. Vienna, Austria R Foundation for Statistical Computing; 2015.
22. McLean B, Coutts A, Kelly V, McGuigan M, Cormack S. Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *Int J Sports Physiol Perform.* 2010;5(3):367-383.
23. Gabbett T. The training-injury prevention paradox: Should athletes be training smarter and harder? *Br J Sports Med.* 2016;50(5):273-280.
24. Hulin B, Gabbett T, Blanch P, Chapman P, Bailey D, Orchard J. Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *Br J Sports Med.* 2013;48:708-712.
25. Hulin B, Gabbett T, Lawson D, Caputi P, Sampson J. The acute: chronic workload ratio predicts injury: High chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med.* 2016;50(4):231-236.
26. Murray N, Gabbett T, Townshend A, Hulin B, McLellan C. Individual and combined effects of acute and chronic running loads on injury risk in elite Australian footballers. *Scand J Med Sci Sports.* 2017;27(9):990-998.
27. Carey D, Blanch P, Ong K, Crossley K, Crow J, Morris M. Training loads and injury risk in Australian football-differing acute: chronic workload ratios influence match injury risk. *Br J Sports Med.* 2017;51(16):1215-1220.
28. Lolli L, Batterham A, Hawkins R, et al. Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. *Br J Sports Med.* 2017;In Press.
29. Buchheit M. Applying the acute:chronic workload ratio in elite football: Worth the effort? *Br J Sports Med.* 2017;51(18):1325-1327.
30. Williams S, West S, Cross M, Stokes K. Better way to determine the acute: chronic workload ratio? *Br J Sports Med.* 2016;51(3):209-210.
31. Cummins C, Delaney J, Thornton H, Duthie G. Training load prior to injury in professional rugby league players. *J Sci Med Sport.* 2017;20 (Suppl 3):S53-54.

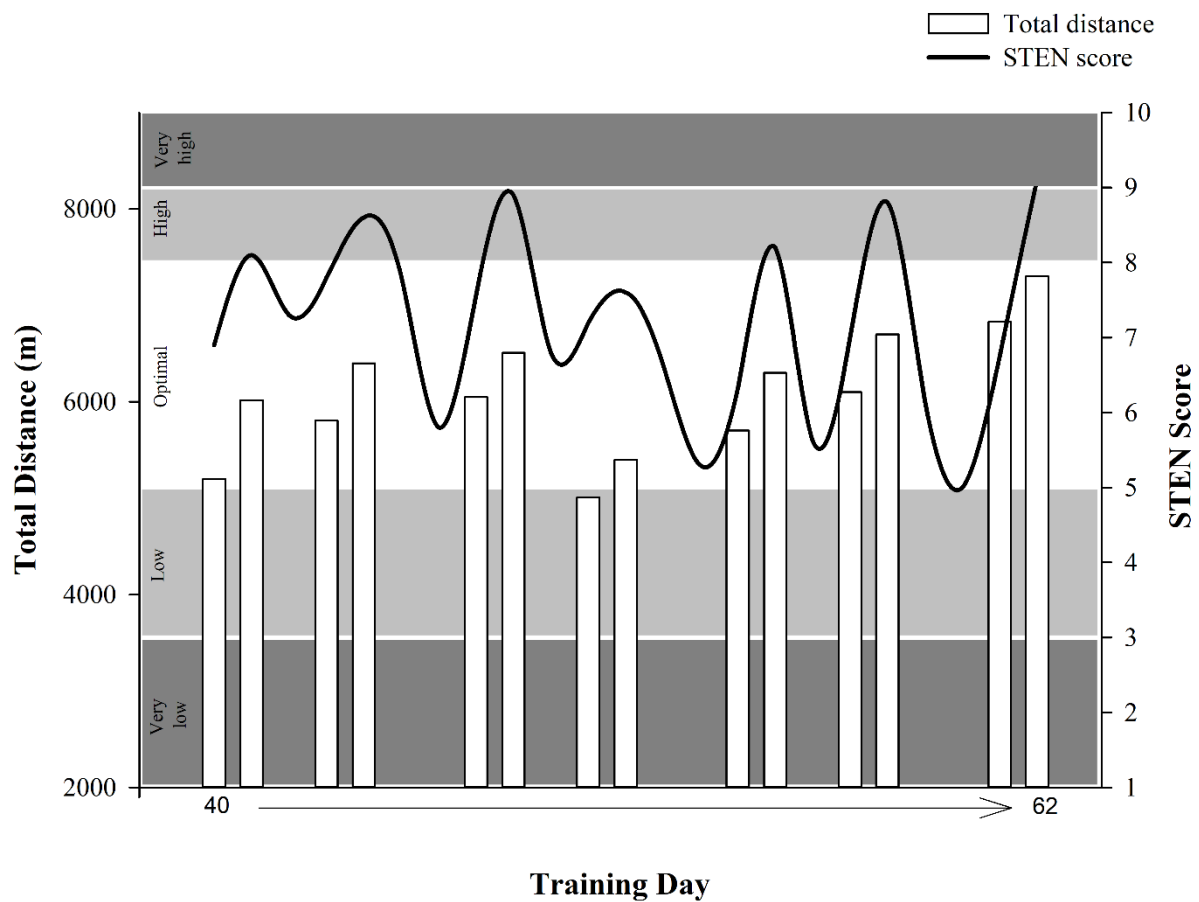
32. Thornton H, Delaney J, Duthie G, Dascombe B. Effects of pre-season training on the sleep characteristics of professional rugby league players. *Int J Sports Physiol Perform.* 2018;13(2):176-182.
33. Murray N, Gabbett T, Townshend A. The use of relative speed zones in Australian football: are we really measuring what we think we are? *Int J Sports Physiol Perform.* 2018;13(4):442-451.
34. Tu YK, Gilthorpe MS. Revisiting the relation between change and initial value: a review and evaluation. *Stat Med.* 2007;26(2):443-457.
35. Windt J, Gabbett TJ. Is it all for naught? What does mathematical coupling mean for acute: chronic workload ratios? *Br J Sports Med.* 2018;In Press.
36. Moore D, McCabe G, Craig B. *Introduction to the practice of statistics.* Sixth edition ed. New York: W. H. Freeman and Company; 2009.
37. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. *Int J Endocrinol Metab.* 2012;10(2):486-489.
38. Robertson S, Bartlett JD, Gatin PB. Red, amber, or green? Athlete monitoring in team sport: The need for decision-support systems. *Int J Sports Physiol Perform.* 2017;12(Suppl 2):S273-S279.
39. Hopkins W, Marshall S, Batterham A, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc.* 2009;41(1):3-13.
40. Buchheit M. The numbers will love you back in return-I promise. *Int J Sports Physiol Perform.* 2016;11(4):551-554.
41. Hopkins W. Measures of reliability in sports medicine and science. *Sports Med.* 2000;30(1):1-15.
42. Pyne D. Interpreting the results of fitness testing. International Science and Football Symposium; 2003; Melbourne.
43. Bernards J, Sato K, Haff G, Bazyler C. Current research and statistical practices in sport science and a need for change. *Sports.* 2017;5(4):87.
44. Rhea M. Determining the magnitude of treatment effects in strength training research through the use of the effect size. *J Strength Cond Res.* 2004;18(4):918-920.
45. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, New Jersey: Erlbaum; 1988.
46. Hopkins W. A spreadsheet for deriving a confidence interval, mechanistic inference and clinical inference from a p value. *Sportscience* 2007; 16-20. Available at: [sportsci.org/2007/wghinf.htm](http://sportsci.org/2007/wghinf.htm), 11.
47. Welsh A, Knight E. "Magnitude-based inference": a statistical review. *Med Sci Sports Exerc.* 2015;47(4):874-884.

48. Sainani K. The Problem with "Magnitude-Based Inference". *Med Sci Sports Exerc.* 2018;In Press.
49. Hopkins W, Batterham A. Error rates, decisive outcomes and publication bias with several inferential methods. *Sports Med.* 2016;46(10):1563-1573.
50. Mengersen K, Drovandi C, Robert C, Pyne D, Gore C. Bayesian estimation of small effects in exercise and sports science. *PloS One.* 2016;11(4):1-23.
51. Hopkins W, Batterham A, Pyne D, Impellizzeri F. Misplaced decimal places. *Scand J Med Sci Sports.* 2011;21(6):867-875.
52. Weissgerber T, Milic N, Winham SJ, Garovic V. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol.* 2015;13(4):e1002128.





**Figure 1.** An outline of the primary steps involved in developing an athlete monitoring system. ACWR = acute to chronic workload ratio, EWMA = exponentially-weighted moving averages, STEN = standard tens score, SWC = smallest worthwhile change, CV = coefficient of variation, MBI = magnitude-based inferences.



**Figure 2.** An example of the use of exponentially-weighted moving averages and an associated standard ten (STEN) score for one athlete. In this example, raw total distance is shown for each training day (starting at training day 40) and the STEN score for the 7-day EWMA is depicted. The STEN scores are categorized into bands, depicted as; 1-3; very low, 3-5; low, 5-8; optimal; 8-9, high; 9-10, very high.

**Table 1.** An example of a simple subjective questionnaire using a 1 to 10-point Likert scale (1 = extremely sore, 10 = no soreness) for soreness responses, and a 5-point descriptive scale for wellbeing. Cells are conditionally formatted according to the raw score for soreness (1 to 2; dark grey, 3 to 4; light grey, and 5 to 10; no shading), wellbeing (1; dark grey, 2; light grey, 3 to 5; no shading) and also the z-score (<-2; dark grey and -1.5 to -1.99; light grey).

Athlete	Perceived Muscle Soreness (1-10)												Wellbeing		
	L/Back	(Z)	Glutes	(Z)	Quads	(Z)	Hams	(Z)	Calves	(Z)	Groin	(Z)	Fatigue	Sleep	Stress
1	5	-1.2	8	-0.7	8	-0.5	5	-1	5	-1.1	8	-0.5	Normal	Good sleep	Very relaxed
2	6	-0.2	5	-0.4	5	-0.4	4	-0.7	5	-0.6	5	-0.7	Normal	Average sleep	Normal
3	4	-1.4	3	-2.1	6	-0.4	2	-2.1	2	-1.5	3	-0.5	Always Tired	Poor Sleep	Feeling Stressed
4	4	-0.2	5	-0.7	5	-0.2	3	-1.3	5	-1.7	5	-0.8	Normal	Good sleep	Normal
5	5	-0.6	6	-1.3	5	-0.2	6	-1.1	6	-0.4	6	-0.7	Normal	Average sleep	Normal
6	6	-0.5	6	-0.3	6	-0.5	5	-0.5	6	-0.5	6	-0.4	Normal	Good sleep	Normal
7	4	-1.6	5	-0.3	7	-0.9	5	-1.1	6	-1	6	-0.1	Normal	Average sleep	Normal
8	5	-0.7	4	-0.7	4	-1.5	3	-2.1	4	-0.6	5	-0.3	More tired than normal	Good sleep	Highly Stressed
9	4	-1.4	4	-2.4	5	-0.9	5	-0.6	6	-0.9	5	-0.7	Normal	Average sleep	Normal
10	7	-0.7	7	-1.2	7	-0.3	7	-0.3	7	-0.3	7	-0.7	Normal	Good sleep	Normal
11	5	-0.1	5	-0.3	5	-0.1	2	-1.8	3	-1.1	5	-0.2	More tired than normal	Good sleep	Normal
12	8	-0.3	7	-1.1	7	-1.1	6	-0.8	7	-0.9	8	-0.8	Normal	Very Refreshed	Normal