Assessing the Measurement Sensitivity and Diagnostic Characteristics of Athlete-Monitoring Tools in National Swimmers

Stephen Crowcroft, Erin McCleave, Katie Slattery, and Aaron J. Coutts

Purpose: To assess measurement sensitivity and diagnostic characteristics of athlete-monitoring tools to identify performance change. **Methods:** Fourteen nationally competitive swimmers (11 male, 3 female; age 21.2 ± 3.2 y) recorded daily monitoring over 15 mo. The self-report group ($n = 7$) reported general health, energy levels, motivation, stress, recovery, soreness, and wellness. The combined group ($n = 7$) recorded sleep quality, perceived fatigue, total quality recovery (TQR), and heart-rate variability. The week-to-week change in mean weekly values was presented as coefficient of variance (CV%). Reliability was assessed on 3 occasions and expressed as the typical error CV%. Week-to-week change was divided by the reliability of each measure to calculate the signal-to-noise ratio. The diagnostic characteristics for both groups were assessed with receiver-operating-curve analysis, where area under the curve (AUC), Youden index, sensitivity, and specificity of measures were reported. A minimum AUC of .70 and lower confidence interval (CI) $>.50$ classified a “good” diagnostic tool to assess performance change. **Results:** Week-to-week variability was greater than reliability for soreness (3.1), general health (3.0), wellness% (2.0), motivation (1.6), sleep (2.6), TQR (1.8), fatigue (1.4), R-R interval (2.5), and LnRMSSD:RR (1.3). Only general health was a “good” diagnostic tool to assess decreased performance (AUC $\sim .70$, 95% CI, .61–.80). **Conclusion:** Many monitoring variables are sensitive to changes in fitness and fatigue. However, no single monitoring variable could discriminate performance change. As such the use of a multidimensional system that may be able to better account for variations in fitness and fatigue should be considered.

Keywords: subjective questionnaires, heart-rate variability, training monitoring, swimming

The primary goal of high-performance coaches and sport science staff is to deliver well-controlled training programs for achieving peak performance. Traditional approaches to training prescription include increases in both training volume and intensity before a taper.¹ However, during such heavy training periods athletes may be at a higher risk of negative training outcomes such as injury, illness, and/or overreaching.² Indeed, these negative outcomes in training have been associated with a lower chance of achieving success in competition.³ Therefore, to better understand if athletes are tolerating life stressors and training demands, athletes are monitored through a number of physiological and subjective measures. While observational studies have shown monitoring tools fluctuate with changes to training load, illness, and overreaching, no studies are yet to examine the sensitivity, specificity, or diagnostic characteristics of these monitoring tools for identifying performance change.^{2,4,5}

In medicine, a common approach used to assess this characteristic is testing a binary outcome, which yields 2 discrete functions to infer an unknown, based on a specific cutoff value (eg, disease or no disease from a specific blood marker). The accuracy for determining these specific cutoff values is assessed using measures of sensitivity (true positive rate, ie, the percentage of accurately diagnosed disease) and specificity (true negative rate, ie, the percentage of patients accurately diagnosed with no disease). These values are then assessed using a receiver operating characteristic (ROC) curve to determine the diagnostic accuracy and optimal cutoff values which

identify a particular outcome.⁶ However, identifying the diagnostic characteristics of monitoring variables on performance change in athletes has yet to be assessed in this manner.

Subjective questionnaires and heart-rate variability (HRV) have been assessed for their practicality when monitoring athletes in the daily training environment.^{5,7} These measures are used to determine an athlete's fitness and/or fatigue status and inform coaches how their athletes are tolerating training demands. Although numerous validated psychometric tools are available, many athletes prefer using abbreviated questionnaires due to their simplicity and shorter time required for completion.^{4,7} Likewise, many physiological measures have been used for monitoring the athlete training response. For example, HRV has been used to monitor athletes' fitness or fatigue status and inform future training decisions.^{5,8} Changes in HRV reflect the autonomic nervous system's control of cardiovascular function from alterations in parasympathetic modulation that may signify an adaptive or maladaptive training state.^{3,8} However, while case studies have identified the positive and negative training responses, whether these variables are sensitive to adjustments in training or performance is unclear. As such, the purpose of this study was, first, to report the week-to-week variability, reliability, and signal-to-noise ratio in common athlete-monitoring tools and, second, to assess the diagnostic characteristics of these tools to identify improvements and decrements in performance.

Methods

Participants

Two groups of 7 ($N = 14$) nationally competitive swimmers (11 male, 3 female; age 21.2 ± 3.2 y, body mass 75.8 ± 6.8 kg, best time in main event as percentage of world record $93.5\% \pm 2.2\%$,

The authors are with the Sport & Exercise Discipline Group, University of Technology Sydney (UTS), Moore Park, NSW, Australia. Crowcroft and Slattery are also with the New South Wales Inst of Sport (NSWIS), Sydney, NSW, Australia. Address author correspondence to Stephen Crowcroft at stephen.crowcroft@uts.edu.au.

mean $\pm$ SD) were monitored daily over 15 months. Each swimmer was provided a verbal and written explanation of the investigation before giving signed informed consent to release their data for the purpose of this research. This study was approved by the human research ethics committee of University of Technology Sydney (Ref No. 2014000842).

Study Design

This longitudinal observational study occurred over a 15-month period. Throughout this period, 2 commonly used monitoring systems were assessed. Monitoring questions from the “self-report” group were part of a subjective athlete self-report monitoring system from the national body (Swimming Australia). The “combined” group provided an alternate system that included both physiological and subjective athlete self-report questions. Two separate groups were used to compare the diagnostic accuracy of a number of commonly used monitoring variables. Swimmers in each group completed a similar training load (average weekly distance—self-report 41 ± 6 km, combined 42 ± 9 km, and duration—self-report 17 ± 3 h, combined 17 ± 2 h). Athletes were asked to complete all monitoring measures daily at a standardized time for each individual and on race days at least 1 hour before competition. Part one of the analysis investigated the test retest reliability (ie, noise) and the week-to-week change (ie, signal) of all measures. The week-to-week change was then divided by test–retest reliability to calculate the signal-to-noise ratio. Part 2 assessed the diagnostic ability of the tools from the ROC analysis, using the area under the curve (AUC) and Youden index as diagnostic classifiers. Daily and 7-day rolling averages of monitoring variables that aligned with races were included in this part of the analysis.

Monitoring Variables

The self-report group completed daily subjective questions as “how you feel today” for general health, motivation, energy levels, stress, and recovery on 1-to-5 scales (1 = very poor to 5 = excellent) and soreness as a 1 to 10 scale which was inverted before analysis (1 = no soreness at all, 10 = unbearable pain). The sum of these measures were also calculated for a total wellness score and expressed as a percentage of highest possible total (wellness%).⁹ These variables were reported on the athlete’s personal smart phone and stored to a cloud-based athlete monitoring software (Smartabase, Fusion Sport, Brisbane, Australia). The 7-day rolling average was included in analysis if at least 3 data points were recorded in the previous 7 days (wellness%_{7d}).

The combined group completed a customized monitoring system which included resting HRV, sleep quality and fatigue rating (1 = much worse than normal, to 5 = much better than normal) and the total quality recovery scale (TQR; 6 to 20 scale). Athletes were asked to report all subjective measures with “how you feel today.” The 7-day rolling averages of both fatigue (fatigue_{7d}) and TQR (TQR_{7d}) were analyzed in the ROC curve if at least 3 data points were available for analysis in the previous 7 days. These variables were reported on the athlete’s personal smart phone and stored to a cloud-based data-management system (Google Docs, USA). On waking each morning HRV was recorded for 6 minutes in a supine position via R-R series using the Polar Team 2 system (1.3.0.3. Polar Electro Oy, Kempele, Finland). Athletes left the heart-rate monitor by their bedside each evening to minimize disturbances in the morning. After recording, data were downloaded and analyzed using Kubios HRV analysis software 2.0 (Biomedical Signal Analysis Group, University of Kuopio, Kuopio, Finland). Time-domain indices selected for analysis as indicators of training adaptation

and fatigue included R-R interval, the log-transformed root-mean-squared sum of the consecutive R-R intervals (LnRMSSD),¹¹ and the LnRMSSD-to-R-R-interval ratio (LnRMSSD:RR). The LnRMSSD and LnRMSSD:RR were assessed on an isolated day for the signal-to-noise ratio and for ROC analysis as a 7-day rolling average if at least 3 points were available for analysis in the previous 7 days (LnRMSSD_{7d}, LnRMSSD:RR_{7d}).¹³

Performance Measures

Race performances included in analysis were electronically timed by official swimming governing bodies (Swimming Australia or Fédération Internationale de Natation [FINA]) at sanctioned events in a standard international 50-m pool. Before racing, swimmers completed a full race preparation warm up. The performance of each swimmer was tracked in their designated events ranging from 50 to 400 m in freestyle, breaststroke, backstroke, butterfly, and individual medley. If multiple races of the same event (eg, heat and final) were recorded on the same day, the fastest time was recorded for analysis. Each swimmer recorded a minimum of five races of the same event throughout the study period. Times for each athlete’s individual events were averaged to give a mean time and the smallest worthwhile change (SWC) was determined as $0.3 \times$ within-swimmer SD of race performance times.¹⁴ Times outside of the SWC were then coded as a dichotomous outcome variable (0 = no change, 1 = change) to assess both improvements (change <SWC) and decrements (change >SWC) in performance in separate analysis.

Statistical Analysis

Week-to-Week Change. All monitoring measures were converted to mean weekly values, analyzed using a customized spreadsheet and reported as the week-to-week change presented as the coefficient of variance and 90% confidence limits (CV%, $\pm 90\%$ CL).¹⁵ A minimum of 3 measures in the week were required for the data to be included in the analysis. To avoid including data at the same time point leading into major competitions in consecutive seasons, only the first 12 months of data were included in the week-to-week analysis.

Test–Retest Reliability. The test–retest reliability was estimated by measurements taken twice within 1 hour for subjective measures and within 15 minutes for HRV, on 3 separate occasions without any physical activity between the 2 collections. These recordings were completed after a day of no training, low intensity (rating of “somewhat hard” or less) and high-intensity training session (rating greater than “hard”) using the session rating of perceived exertion.¹⁶ For both measures, athletes were asked to record their answers with “how they feel today” at the time of entry without consciously attempting to repeat previous results. Data were then entered into a customized spreadsheet as single-day variables and presented as the test–retest technical error (TE) CV%.¹⁵

Signal-to-Noise Ratio. The signal-to-noise ratio was calculated by dividing the week-to-week change by the reliability of each measure. The signal-to-noise ratio was classified as good if >1 , Ok if ≈ 1 , and poor if <1 .¹⁷

Data Normalization. After the signal-to-noise ratio analysis, monitoring variables were normalized using Z-scores to standardize athlete values. Athletes’ Z-scores were calculated by the daily measure subtracted from the individual’s mean value from all measures collected through the study and divided by the SD of the mean. All Z-scores were then aligned with dichotomous outcomes (0 = no change, 1 = change) for improvements and decrements in performance from

separate analysis. Z-scores were used in this analysis to provide a standardized value and cutoff score (meaningful Z-score) to discriminate performance change relevant for each individual.

Receiver-Operating-Characteristic Curves. Variables with a “good” signal-to-noise ratio and the 7-day rolling average of measures were assessed as Z-scores to examine the diagnostic capability to discriminate a change in performance. Sensitivity versus specificity analysis was performed using ROC curves to assess if a monitoring variable can diagnose an improvement (change < SWC) or decrease in performance (change > SWC) in separate analysis. An ROC curve plots the true positive rate (sensitivity) against the true negative rate (specificity) to produce an AUC. An AUC of 1.00 (100%) represents perfect discriminant power, where .50 (50%) would represent no discriminatory power. An AUC >.70 and the lower CI >.50 was classified as a “good” benchmark.¹⁸ All ROC curve results were presented as AUC $\pm$ 95% CI (6). The Youden index was calculated (Youden index = sensitivity + specificity - 1) from all ROC curve plots (sensitivity and specificity) to determine the point where the Youden index was optimized and considered the score at which a cutoff value presented as a Z-score from each monitoring variable might be acceptable to discriminate a change in performance. The maximum Youden index of 1 would suggest perfect discriminatory value, whilst a score of 0 would reflect no diagnostic value.¹⁹ All ROC analysis was performed using SPSS (version 21, IBM Co, New York, USA).

Results

Performance

The total number of races included in the analysis, along with the number of improvements (change < SWC) and decrements (change > SWC) in performance are reported in Tables 1 and 2, respectively. The

self-report group had a total of 143 performances included in analysis, recording 27 ± 8 races/athlete. The combined group had a total of 143 performances included in analysis, recording 27 ± 6 races/athlete.

Signal-to-Noise Ratio

The week-to-week CV%, reliability, and signal-to-noise ratio are reported in Table 3. For both groups a total of 280 training weeks (40 wk/athlete) were analyzed for the week-to-week CV% respectively. For both groups, 4 weeks were excluded from analysis due to no training.

The seven athletes in the self-report group had a total of 252 athlete training weeks included in analysis. The combined group had a combined 243 training weeks included for the subjective measures (sleep, fatigue, and TQR) and 226 training weeks for HRV across all 7 athletes.

Of all measures assessed there was a “good” signal-to-noise ratio for the following variables in the self-report group: soreness (3.1), general health (3.0), wellness% (2.0), and motivation (1.6). The combined group had a “good” signal-to-noise ratio for sleep (2.6), TQR (1.8), fatigue (1.4), R-R interval (2.5), and LnRMSSD:RR (1.3).

Receiver-Operating-Characteristic Curves

For improvements and decrements in performance the AUC, Z-score with optimal sensitivity, specificity and Youden index are reported in Table 1 and 2 respectively. Only general health from the self-report group met the criteria as a “good” measure to diagnose decrements in performance (AUC -.70, 95% CI .61-.80). Furthermore, an optimal cutoff Z-score of <0.2 for general health from the self-report group was identified to discriminate decreases in performance (Youden index -0.34, sensitivity -62%, specificity -73%). To discriminate positive changes in performance the optimal cutoff Z-score of >0.9

Table 1 Diagnostic Characteristics of Monitoring Variables for Improvements in Performance

Group	Variable	Improved performances, n	Total performances, n	Area under curve (95% CI)	Z-score	Sensitivity	Specificity	Youden index
Self-report	Performance	52	143					
	General health	45	124	.67 (.57-.76)	>0.2	78%	51%	0.28
	Motivation	45	124	.58 (.48-.68)	>0.4	56%	62%	0.18
	Soreness	45	124	.61 (.51-.71)	>1.0	51%	72%	0.23
	Wellness%	45	124	.62 (.51-.72)	>0.6	69%	57%	0.26
	Wellness% _{7d}	50	136	.64 (.54-.74)	>1.0	54%	76%	0.30
Combined	Performance	66	143					
	Sleep	66	142	.52 (.43-.62)	>0.1	50%	59%	0.09
	TQR	66	142	.62 (.53-.72)	>0.4	59%	65%	0.24
	TQR _{7d}	67	143	.64 (.54-.73)	>0.9	60%	75%	0.35
	Fatigue	66	142	.65 (.56-.74)	>0.1	71%	54%	0.25
	Fatigue _{7d}	67	143	.66 (.56-.75)	>0.7	61%	71%	0.32
	RR interval	46	99	.65 (.54-.76)	<-0.7	54%	79%	0.34
	LnRMSSD:RR _{7d}	55	120	.62 (.52-.72)	>0.4	60%	66%	0.26

Abbreviations: CI, 95% confidence intervals; wellness%, a sum of all subjective measures expressed as a percentage of highest possible total; wellness%_{7d}, the 7-d rolling average of wellness%; TQR, total quality recovery scale; TQR_{7d}, The 7-d rolling average of the TQR; fatigue_{7d}, the 7-d average of the fatigue; RR interval, the average R-R interval from morning resting heart rate; LnRMSSD:RR_{7d}, the 7-d average of the LnRMSSD:R-R interval ratio.

Table 2 Diagnostic Characteristics of Monitoring Variables for Decrements in Performance

Group	Variable	Decreased performances, n	Total performances, n	Area under curve (95% CI)	Z-score	Sensitivity	Specificity	Youden index
Self-report	Performance	48	143					
	General health	47	124	.70 (.61–.80)	<0.2	62%	73%	0.34
	Motivation	47	124	.63 (.52–.73)	<0.2	66%	58%	0.24
	Soreness	47	124	.59 (.49–.70)	<-0.1	32%	87%	0.19
	Wellness%	47	124	.63 (.53–.73)	<1.3	89%	33%	0.22
	Wellness% _{7d}	44	136	.64 (.54–.73)	<1.0	84%	45%	0.29
Combined	Performance	54	143					
	Sleep	54	142	.56 (.47–.66)	<-0.2	37%	74%	0.11
	TQR	54	142	.62 (.52–.71)	<0.1	52%	68%	0.20
	TQR _{7d}	54	143	.62 (.53–.72)	<0.7	67%	61%	0.27
	Fatigue	54	142	.61 (.52–.70)	<0.1	56%	66%	0.22
	Fatigue _{7d}	54	143	.63 (.53–.72)	<1.1	89%	38%	0.27
	RR interval	38	99	.59 (.48–.70)	>-0.7	79%	46%	0.25
	LnRMSSD:RR _{7d}	46	120	.59 (.48–.70)	<0.7	80%	39%	0.19

Abbreviations: CI, 95% confidence intervals; wellness%, a sum of all subjective measures expressed as a percentage of highest possible total; wellness%_{7d}, the 7-d rolling average of wellness%; TQR, total quality recovery scale; TQR_{7d}, The 7-d rolling average of the TQR; fatigue_{7d}, the 7-d average of the fatigue; RR interval, the average R-R interval from morning resting heart rate; LnRMSSD:RR_{7d}, the 7-d average of the LnRMSSD:R-R interval ratio.

Note: **Bold text** highlights diagnostic criteria of area under the curve >.70 and lower CI >.50 to be classified as a “good” discriminatory monitoring variable.

Table 3 Week-to-Week Variation, Test–Retest, and Signal-to-Noise Ratio of Monitoring Variables

Monitoring variable	Week-to-week change (90% CL)	Test–retest (90% CL)	Signal-to-noise ratio	Rating
Soreness	8.0 (7.3 – 8.9)	2.6 (2.0–3.8)	3.1	Good
General health	10.6 (9.6 – 11.8)	3.5 (2.7–5.2)	3.0	Good
Wellness%	4.9 (4.5 – 5.5)	2.5 (2.0–3.7)	2.0	Good
Motivation	10.0 (9.1 – 11.1)	6.2 (4.8–9.3)	1.6	Good
Energy levels	10.6 (9.7 – 11.9)	10.4 (7.9–15.5)	1.0	Ok
Recovery	8.0 (7.3 – 8.9)	8.4 (6.4–12.6)	1.0	Ok
Stress	7.9 (7.2 – 8.8)	9.3 (7.1–13.9)	0.8	Poor
Sleep	11.5 (10.5 – 12.9)	4.5 (3.5–6.7)	2.6	Good
Total quality recovery scale	6.8 (6.2 – 7.6)	3.8 (2.9–5.5)	1.8	Good
Fatigue	14.0 (12.6 – 15.7)	10.0 (7.6–15.0)	1.4	Good
R-R interval	8.9 (8.0 – 10.0)	3.5 (2.7–5.1)	2.5	Good
LnRMSSD:RR ratio	10.1 (9.1 – 11.4)	7.7 (5.9–11.4)	1.3	Good
LnRMSSD	5.9 (5.4 – 6.6)	6.3 (4.8–9.4)	0.9	Poor

Abbreviations: 90% CL, 90% confidence limits; signal-to-noise ratio, calculated from dividing the week-to-week change by the test–retest coefficient of variance; rating, ratings of the signal-to-noise ratio good if >1, Ok if ~1, and poor if <1; wellness%, sum of variables from a total wellness score expressed as a percentage of highest possible total; LnRMSSD:RR, the LnRMSSD-to-R-R-interval ratio; LnRMSSD, the log-transformed root-mean-squared sum of the consecutive R-R intervals.

was identified for TQR_{7d} (Youden Index –0.35, sensitivity –60%, specificity –75%) from the combined group.

Discussion

It appears that several monitoring variables have poor discriminatory ability for assessing worthwhile variations in performance. However, there was a good signal-to-noise ratio for many measures

including soreness, general health, wellness%, motivation, sleep, TQR, fatigue, R-R interval, and the LnRMSSD:RR. As such, changes in fitness and fatigue of well-trained swimmers can be tracked with self-reported and HRV indices but these measures are poor discriminators of meaningful changes in performance.

The first part of this study assessed the test–retest reliability, week-to-week variation, and signal-to-noise ratio of commonly used monitoring tools. Many of the monitoring tools were sensitive

(ie, typical weekly changes > test–retest error) to changes in the swimmer’s training programs. The magnitude of typical weekly changes of general health, soreness, wellness%, motivation, sleep, TQR, fatigue, R-R interval, and the LnRMSSD:RR ratio were between 1.3 to 3.1 times the test–retest error. The variations in these subjective questionnaires are similar to previous studies in well-trained swimmers.^{20,21} For example, over a 6 month training period Morgan et al²⁰ identified that fluctuations in mood for the 41 well-trained swimmers was associated with changes in training load. Noticeable increases in mood disturbances from the profile of mood states during peak training periods returned to baseline values throughout a taper. Similarly, during a deliberate 10-day increase in training load, self-reported general well-being, and muscle soreness reflected individual tolerance to training demands in 12 swimmers.²¹ Collectively, these findings demonstrate the ability of wellbeing measures to assess fitness and fatigue in swimmers. Moreover, applying a signal-to-noise ratio should assist practitioners to better quantify changes in fitness and fatigue measures.

The typical week-to-week variation of subjective questionnaires in the current study was noticeably lower than similar questionnaires in team sports.^{22,23} The reliability CV% of subjective questionnaires over 2 similar training weeks ranged from 12% on game day to 32% the day immediately after the game in Australian football players.²² Similar, larger variations (7.1 CV%) and poor sensitivity of wellness questionnaires was reported in rugby league players.¹⁷ However, sensitivity was determined from the between-days reliability as a $CV\% \times 0.2$ to calculate the SWC without training between tests. The larger variations compared with the results in the current study (4.9% week-to-week and 2.5% test–retest) could be attributed to the differences in methodology calculating typical error and also reflect the differences in the training and match demands of team sport athletes compared with swimmers. Swimming training and competition is highly controlled, whereas team-sport athletes have other external factors including weekly fixtures, tactical changes, and collisions that may further affect the variability in their subjective responses.

Similar to the wellness questionnaires, the signal (ie, training-induced changes) was greater than the test–retest noise for both the R-R interval and LnRMSSD:RR. Notably both the reliability and weekly CV% of HRV from this study were also lower than previous reports. For example this study reported CV% for R-R interval and LnRMSSD (3.5% and 6.3%) were lower than between-days CV% in team-sport athletes (11.1% and 12.3%, respectively).¹¹ Accounting for methodological differences, this variance may be explained from comparing moderately trained (5–6 h/wk) to the highly trained participants (18–22 h/wk) in this study. The high sensitivity of HRV in the present investigation provides evidence that in swimmers, R-R interval and LnRMSSD:RR may be useful to guide decision making around training prescription. Several studies have shown that endurance-training outcomes can be improved through modifying training prescription based on daily changes of HRV in recreational athletes (24–26). Surprisingly, however, no studies have described the typical daily variability in these measures compared with the same-day test–retest in highly trained athletes. Taken collectively, HRV and subjective measures are sensitive to training induced changes in fitness and fatigue.

The poor diagnostic characteristics of monitoring variables raise questions of their appropriateness to identify performance change. Only general health (AUC = .70, CI .61–.80) in the self-report group fit the benchmark criteria of AUC >.70 and lower CI >.50 to be classified as a “good” discriminatory variable for decreased performance. These findings are similar to Saw et al,⁴

whose systematic review indicated that subjective questionnaires are responsive to training load but were unable to discriminate performance change in a variety of athletes. However, fatigue could explain up to 53% of the variance in performance of elite swimmers and by combining sleep, stress, fatigue, and muscle soreness this explanation has been increased up to 72%.²⁷ While selected monitoring measures can assess performance after a taper, the accuracy of a single measure to predict performance change throughout a season with a larger sample size is not as strong.

The Youden index assesses the balance between sensitivity and specificity to quantify cutoff values identifying if an outcome is likely to occur. The poor discriminatory function of the monitoring variables is highlighted with the highest value for improved performance being TQR_{7d} (Youden index = 0.35, Z-score >0.9, sensitivity = 60%, specificity = 75%) and decreased performance from general health (Youden index = 0.34, Z-score <0.2, sensitivity = 62%, specificity = 73%). However, as the current study is the first to use the Youden index to assess athlete-monitoring tools, an acceptable threshold and interpretation of sensitivity and specificity to accurately identify changes in athletic performance is still relatively unknown. Indeed, little evidence is currently available to support how specific cutoff values are implemented. Reports from high performance sports on athlete monitoring suggest there is a strong reliance on the visual analysis of trends for when to modify training.⁷ While the Youden index in the present investigation showed a weak association to changes in performance, future use of this data analysis technique may help provide meaningful cutoff values to enhance the ability of monitoring tools to assess performance change.

Assessing performance accurately can be problematic when monitoring variables are examined in isolation. An individual’s capacity to perform on a given day is dependent on a multitude of factors ranging from physiological, nutritional, and/or psychological. While subjective measures may be able to detect fatigue, they provided little understanding of an athlete’s performance capacity throughout a season. It is also difficult to classify an improvement or decrement in the performance of elite swimmers. In this investigation, SWC was used without consideration of the TE of race performance. This was due to the difficulty in assessing a true TE from 2 maximal race performances. Nonetheless, it should be acknowledged that this may influence the number of calculated improvements and decrements in performance, in turn affecting the ROC analysis and results. Moreover, the use of Z-scores to normalize data in this study may have weakened the intraindividual diagnostic accuracy. Furthermore, a multidimensional psychobiological approach to monitoring may provide more useful information to coaches. Through the longitudinal analysis of the 2 separate training groups, the multitude of variables assessed could assist sport scientists in the selection of useful monitoring tools for a more holistic and effective assessment of an athlete’s fitness and fatigue.

Practical Applications

Monitoring the general health of athletes was the only variable that met the criteria to assess decrements in performance. We recommend from the results of both the AUC and the Youden index that caution be taken with the interpretation of any single monitoring tool to assess performance change. Further, it is appropriate that monitoring tools are validated with each sport. The smaller CV% in swimmers compared with team sports suggests that even subtle variations in subjective and HRV measures may be important in these athletes.

Conclusion

The lack of discriminatory ability of any single monitoring variable brings into question the utility of this approach to accurately assess performance change. The “good” signal-to-noise ratio of the numerous monitoring tools assessed shows the variables potential to monitor athletes’ fitness and fatigue. However, both the type of athlete and sport should be considered in establishing a monitoring system. A multidimensional monitoring system to account for variations in fatigue and an athlete’s performance capacity should therefore be considered.

References

1. Kiely J. Periodization paradigms in the 21st century: evidence-led or tradition-driven? *Int J Sports Physiol Perform*. 2012;7(3):242–250. [PubMed doi:10.1123/ijspp.7.3.242](#)
2. Meeusen R, Duclos M, Foster C, et al. Prevention, diagnosis and treatment of the overtraining syndrome: joint consensus statement of the European College of Sport Science (ECSS) and the American College of Sports Medicine (ACSM). *Eur J Sport Sci*. 2013;13(1):1–24. [doi:10.1080/17461391.2012.730061](#)
3. Raysmith BP, Drew MK. Performance success or failure is influenced by weeks lost to injury and illness in elite Australian track and field athletes: a 5-year prospective study. *J Sci Med Sport*. 2016; in press. [PubMed doi:10.1016/j.jsams.2015.12.515](#)
4. Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *Br J Sports Med*. 2016;50(5):281–291. [PubMed doi:10.1136/bjsports-2015-094758](#)
5. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? *Front Physiol*. 2014;5(27):73. [PubMed](#)
6. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861–874. [doi:10.1016/j.patrec.2005.10.010](#)
7. Taylor K, Chapman DW, Cronin JB, Newton MJ, Gill N. Fatigue monitoring in high performance sport: a survey of current trends. *J Austr Strength Cond*. 2012;20(1):12–23.
8. Bellenger CR, Fuller JT, Thomson RL, Davison K, Robertson EY, Buckley JD. Monitoring athletic training status through autonomic heart rate regulation: a systematic review and meta-analysis. *Sports Med*. 2016;46(10):1461–1486. [PubMed](#)
9. Hooper SL, Mackinnon LT. Monitoring overtraining in athletes: recommendations. *Sports Med*. 1995;20(5):321–327. [PubMed doi:10.2165/00007256-199520050-00003](#)
10. Kenttä G, Hassmén P. Overtraining and recovery: a conceptual model. *Sports Med*. 1998;26(1):1–16. [PubMed doi:10.2165/00007256-199826010-00001](#)
11. Al Haddad H, Laursen PB, Chollet D, Ahmaidi S, Buchheit M. Reliability of resting and post exercise heart rate measures. *Int J Sports Med*. 2011;32(8):598–605. [PubMed doi:10.1055/s-0031-1275356](#)
12. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. *Sports Med*. 2013;43(9):773–781. [PubMed doi:10.1007/s40279-013-0071-8](#)
13. Plews DJ, Laursen PB, Le Meur Y, Hausswirth C, Kilding AE, Buchheit M. Monitoring training with heart rate variability: how much compliance is needed for valid assessment? *Int J Sports Physiol Perform*. 2014;9(5):783–790. [PubMed doi:10.1123/ijspp.2013-0455](#)
14. Hopkins WG, Hawley JA, Burke LM. Design and analysis of research on sport performance enhancement. *Med Sci Sports Exerc*. 1999;31(3):472–485. [PubMed doi:10.1097/00005768-199903000-00018](#)
15. Hopkins WG. Analysis of reliability with a spreadsheet. *Sportscience*. <http://sportsci.org/resource/stats/xrely.xls>2010
16. Seiler KS, Kjerland Ø. Quantifying training intensity distribution in elite endurance athletes: is there evidence for an “optimal” distribution? *Scand J Med Sci Sports*. 2006;16(1):49–56. [PubMed doi:10.1111/j.1600-0838.2004.00418.x](#)
17. Roe G, Darrall-Jones J, Till K, et al. Between-day reliability and sensitivity of common fatigue measures in rugby players. *Int J Sports Physiol Perform*. 2016;11(5):581–586. [PubMed doi:10.1123/ijspp.2015-0413](#)
18. Menaspà P, Sassi A, Impellizzeri FM. Aerobic fitness variables do not predict the professional career of young cyclists. *Med Sci Sports Exerc*. 2010;42(4):805–812. [PubMed doi:10.1249/MSS.0b013e3181ba99bc](#)
19. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005;16(1):73–81. [PubMed doi:10.1097/01.ede.0000147512.81966.ba](#)
20. Morgan WP, Brown DR, Raglin JS, O’Connor PJ, Ellickson KA. Psychological monitoring of overtraining and staleness. *Br J Sports Med*. 1987;21(3):107–114. [PubMed doi:10.1136/bjism.21.3.107](#)
21. Morgan WP, Costill DL, Flynn MG, Raglin JS, O’Connor PJ. Mood disturbance following increased training in swimmers. *Med Sci Sports Exerc*. 1988;20(4):408–414. [PubMed doi:10.1249/00005768-198808000-00014](#)
22. Gastin PB, Meyer D, Robinson D. Perceptions of wellness to monitor adaptive responses to training and competition in elite Australian football. *J Strength Cond Res*. 2013;27(9):2518–2526. [PubMed doi:10.1519/JSC.0b013e31827fd600](#)
23. Montgomery PG, Hopkins W. The effects of game and training loads on perceptual responses of muscle soreness in Australian football. *Int J Sports Physiol Perform*. 2013;8(3):312–318. [PubMed doi:10.1123/ijspp.8.3.312](#)
24. Vesterinen V, Nummela A, Heikura I, et al. Individual endurance training prescription with heart rate variability. *Med Sci Sports Exerc*. 2016;48(7):1347–1354. [PubMed doi:10.1249/MSS.0000000000000910](#)
25. Kiviniemi AM, Hautala AJ, Kinnunen H, et al. Daily exercise prescription on the basis of HR variability among men and women. *Med Sci Sports Exerc*. 2010;42(7):1355–1363. [PubMed doi:10.1249/MSS.0b013e3181cd5f39](#)
26. Kiviniemi AM, Hautala AJ, Kinnunen H, Tulppo MP. Endurance training guided individually by daily heart rate variability measurements. *Eur J Appl Physiol*. 2007;101(6):743–751. [PubMed doi:10.1007/s00421-007-0552-2](#)
27. Hooper SL, Mackinnon LT, Howard A, Gordon RD, Bachmann AW. Markers for monitoring overtraining and recovery. *Med Sci Sports Exerc*. 1995;27(1):106–112. [PubMed doi:10.1249/00005768-199501000-00019](#)