Is heart rate a convenient tool to monitor over-reaching? A systematic review of the literature

L Bosquet,1,2 S Merkari,1 D Arvisais,3 A E Aubert4

ABSTRACT

Objective: A meta-analysis was conducted on the effect of overload training on resting HR, submaximal and maximal exercise HR (HR), and heart rate variability (HRV), to determine whether these measures can be used as valid markers of over-reaching.

Methods: Six databases were searched using relevant terms and strategies. Criteria for study inclusion were: participants had to be competitive athletes, an increased training load intervention had to be used, and all necessary data to calculate effect sizes had to be available. An arbitrary limit of 2 weeks was chosen to make the distinction between short-term and long-term interventions. Dependent variables were HR and HRV (during supine rest). Standardised mean differences (SMD) in HR or HRV before and after interventions were calculated, and weighted according to the within-group heterogeneity to develop an overall effect.

Results: In these competitive athletes, short-term interventions resulted in a moderate increase in both resting HR (SMD = 0.55; p = 0.01) and low frequency/high frequency ratio (SMD = 0.52; p = 0.02), and a moderate decrease in maximal HR (SMD = −0.75; p = 0.01). Long-term interventions resulted in a small decrease in HR during submaximal (SMD = −0.38; p = 0.006) and maximal exercise (SMD = −0.33; p = 0.007), without alteration of resting values.

Conclusion: The small to moderate amplitude of these alterations limits their clinical usefulness, as expected differences may fall within the day-to-day variability of these markers. Consequently, correct interpretation of HR or HRV fluctuations during the training process requires the comparison with other signs and symptoms of over-reaching to be meaningful.

Peak performance in sport requires training loads that will occasionally push the adaptation capabilities of the human body to their limits. If significant improvements are expected when sufficient recovery is allowed before competition,1,2 athletes may also experience unexplained under-performance.3,4 Various terms have been used in the scientific literature over the past decades to describe this disruption in the training adaptation process. In an effort to clarify the terminology, the Task Force of the European College of Sports Science distinguished functional over-reaching (FOR), non-functional over-reaching (NFOR) and overtraining syndrome (OTS) to mark out the continuum that leads to long-term underperformance.5

Over-reaching refers to an accumulation of training and/or non-training stress resulting in short-term decrease in performance capacity with or without related physiological and psychological signs and symptoms of overtraining, in which restoration of performance capacity may take several days to several weeks.6–7 FOR is occasionally used by athletes during a typical training cycle. An improvement in performance capacity above what would be achievable after a normal training cycle is expected if the athlete is allowed sufficient recovery to permit the occurrence of the super-compensation effect. As suggested by Meeusen et al.,8 when intensified training continues, the athlete can develop a state of extreme over-reaching (ie, NFOR), that will lead to stagnation or a decrease in performance capacity, which will not recover for several weeks.

OTS is usually defined as an accumulation of training and/or non-training stress resulting in long-term decrease in performance capacity with or without related physiological and psychological signs and symptoms of overtraining, in which restoration of performance capacity may take several weeks to several months.5–7 OTS represents the endpoint of the overtraining continuum and is relatively rare, at least in the scientific literature. With the exception of some anecdotal reports,8–10 most of the research that has been conducted in this area has investigated FOR and/or NFOR.8–11

Although the consequences of FOR and NFOR on an athlete’s career can differ dramatically, simply in the time required to recover fully, the limit between these two states is very narrow. Thus there is a need for athletes, coaches and sport scientists to have valid markers of FOR at their disposal, so that appropriate rest can be provided before the development of NFOR and eventually OTS.

Heart rate (HR) is probably one of the most accessible physiological measures in sports medicine. Owing to the development of HR monitors, HR is now widely used in the monitoring of training load and performance capacity.12 According to the model of Rosenblueth and Simeone,13 HR and its modulation are primarily determined by the inotropic and the chronotropic effect of both branches of the autonomic nervous system (ANS) on the myocardium and the sinus node. Increased parasympathetic nervous activity slows HR, whereas increased sympathetic nervous activity accelerates HR. ANS also fulfils a pivotal role in stress tolerance. The scientific and clinical literature suggests that FOR, NFOR and OTS, as well as severe depression, are concomitant with dysfunction of the hypothalamic–pituitary–adrenal axis.14–15 Consequently, negative adaptation to training stress potentially involves the ANS, and may result in a concomitant alteration in HR. This is probably the reason why HR has long been
suggested as a potential marker of FOR, NFOR and OTS.\textsuperscript{16–23} Heart rate variability (HRV) has also been claimed to be a promising tool in this respect.\textsuperscript{24–27} HRV is an index of interbeat intervals; the higher the HRV, the higher the cardiovascular autonomic responsiveness. The specific activity of each branch of the ANS can be estimated using frequency domain analysis. The spectral energy of the high frequency (HF) band (0.15–0.40 Hz) is under parasympathetic control, whereas that of the low frequency (LF) band (0.04–0.15 Hz) reflects both sympathetic and parasympathetic control.\textsuperscript{28} Consequently, the LF/HF ratio is an index of the sympathovagal balance, and is considered with other HRV indexes to be a non-invasive measure of the autonomic cardiovascular control. Although the rationale for using HR and HRV as markers of FOR, NFOR and OTS is sound,\textsuperscript{24–30} recent reviews have underlined the lack of consistency in overload-induced alterations of HR, and the lack of experimental data to support the use of HRV in the follow-up of elite athletes.\textsuperscript{31–34}

Hence, although numerous studies have been performed to explore the relationship between increased training load and HR or HRV, these have used relatively small samples, and thus valuable information could be gained from a systematic review of the literature. The purpose of this investigation was to quantify the effect of increased training load on HR or HRV, to determine whether they can be used as valid markers of FOR, NFOR and OTS. As the duration of increased training load has been suggested to play a key role in the development of NFOR or OTS,\textsuperscript{35} it has been used as a moderator variable. An arbitrary limit of 2 weeks was chosen to make the distinction between short-term and long-term interventions. Dependant variables were resting HR, HR during submaximal or maximal exercise, post-exercise HR recovery and HRV.

**METHODS**

**Literature search**

The databases Embase, Kinpubs, Physical Education Index, PubMed, SportDiscus and Web of Science were searched using the terms "(over-reaching OR overtraining OR overload OR staleness) and (HR OR HR variability)". The reference lists of the articles obtained were searched manually to obtain further studies not identified electronically. This led to the identification of 120 potential studies for inclusion in the analysis.

**Criteria**

The criteria for study inclusion were that participants had to be competitive athletes, the study must have used an increased training load intervention, and the study had to include all necessary data to calculate effect sizes (ie, number of subjects, mean and SD). Additional criteria were used for the inclusion of studies measuring HRV. The sampling rate for the R-wave detection had to be at least 250 Hz, the authors must have provided reasonable assurance of correct processing of the RR interval data, and the spectral power of HRV must have been derived from fast Fourier transformation or autoregressive modelling. It is worth noting that studies reporting only the SD of normal to normal intervals (SDNN) as a measure of HRV were also included in this metaanalysis. As it is the square root of variance, SDNN is mathematically equal to the total power of spectral analysis.\textsuperscript{36} The single criterion for study exclusion was previous publication of the set of data in another article that has already been included in the analysis. Hence, a total of 34 studies were included in the final analysis.\textsuperscript{14–17, 30–61}

**Coding for the studies**

Each study was read and coded independently by two investigators using HR and duration of the overload intervention as moderator variables. An interval scale was used for the coding of HR, and a nominal scale was used for the coding of the duration (\(<2 \text{ weeks}, \geq 2 \text{ weeks}\)). The coding agreement between investigators was determined by dividing the variables with the same codes by the total number of variables. A mean agreement of 0.90 is generally accepted as an appropriate level of reliability in the coding procedure\textsuperscript{62}, and mean agreement was 0.943 in our study. Each coding difference was scrutinised by both investigators together, and resolved before the analysis.

**Statistical heterogeneity**

Statistical heterogeneity refers to the percentage of the variability between studies that is due to clinical and methodological heterogeneity rather than sampling error. In our study, statistical heterogeneity was assessed by the I\textsuperscript{2} statistic, computed as follows:

\[
I^2 = \frac{Q-(k-1)}{Q} \times 100
\]

where Q is the \(\chi^2\) value in a test for heterogeneity and k is the number of studies. A value >50% is considered to be substantial heterogeneity, and may provoke further investigation through subgroup analysis of moderator variables, even if the overall effect is considered non-significant.\textsuperscript{63}

**Statistical analysis**

The standardised mean differences in HR before and after over-reaching were calculated according to the equation below and weighted according to the within-group heterogeneity to develop an overall effect

\[
SMD = \frac{M_2-M_1}{SD_{pooled}}
\]

where SMD is the standardised mean difference, M\textsubscript{1} and M\textsubscript{2} are the mean of the first and the second trial and SD\textsubscript{pooled} is the pooled standard deviation, calculated as follows:

\[
SD_{pooled} = \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{(n_1+n_2-2)}}
\]

where \(S_1^2\) and \(S_2^2\) are the variance of the first and the second trial, and n is the number of subjects.

Statistical significance was set at \(p<0.05\) for all analyses. The scale proposed by Cohen\textsuperscript{64} was used for interpretation. The magnitude of the difference was considered to be small (0.2), moderate (0.5), or large (0.8). All calculations were made with Review Manager V.4.2.8 (The Nordic Cochrane Center, The Cochrane Collaboration, Copenhagen, Denmark).

**RESULTS**

In these competitive athletes, over-reaching resulted in a small decrease in HR measured during submaximal (SMD = −0.27; \(Z = 2.22, p = 0.03\)) and maximal (SMD = −0.38; \(Z = 3.45, p = 0.0003\)) exercise, together with a small increase in the cardiovascular autonomic balance at rest (LF/HF: SDM = 0.41; \(Z = 2.29, p = 0.01\)); there were no data available for post-exercise HR recovery (table 1).
With the exception of maximal HR, which was decreased under all conditions (p < 0.01), there seemed to be a duration-dependent pattern in overload-induced alterations (table 2).

Interventions of ≤2 weeks in duration resulted in a moderate increase in both the cardiovascular autonomic balance (SMD = 0.52; Z = 2.28, p = 0.02) and resting HR (SMD = 0.55; Z = 2.48, p = 0.01), whereas submaximal HR remained unaltered (NS). Conversely, overload interventions lasting >2 weeks resulted in a small decrease in HR during both submaximal (SMD = -0.58; Z = 2.74, p = 0.006) and maximal exercise (SMD = 0.33; Z = 2.69, p = 0.007), without alteration in resting values (NS). Weighted mean differences for resting and maximal HR are reported in table 3.

The 95% CI for the resting tachycardia found after <2 weeks of overload intervention ranged from 2.2 to 6.8 beats/minute. The 95% CI for the bradycardia found in maximal HR ranged from 2.7 to 12.2 beats/minute after an overload intervention of >2 weeks.

A small decrease in HR during both submaximal and maximal exercise after overtraining (p = 0.01) suggests that an increase in resting HR may be used as a valid sign of short-term fatigue (possibly FOR), but not long-term fatigue (possibly NFOR or OTS).

### Submaximal heart rate

Kuipers and Keizer hypothesised that the motor units normally recruited and involved during a particular type and intensity of exercise will be prematurely fatigued during FOR, NFOR and OTS. The possible consequence is an increase in the nervous stimulation of the motor units involved and/or the recruitment of additional motor units, both resulting in an increased HR at a given submaximal workload. Although Fry et al agreed with this possibility, Lehmann et al anticipated a reduction in HR instead, attributed to a decreased sensitivity to catecholamine. In accordance with this second hypothesis, we found a small decrease in HR at the same submaximal workload (p = 0.03), which was particularly evident after long-term interventions (p = 0.006). The absence of alteration after short-term increase in training load (p = 0.91) suggests that this marker cannot be considered as a valid sign of short-term fatigue, but it may be useful in the detection of long-term fatigue.

### Maximal heart rate

Maximal HR is commonly used in the monitoring of exercise intensity in the field. Its use as a marker of FOR, NFOR and OTS is more recent, as no mention was made about its potential usefulness as a sign of FOR, NFOR and OTS. However, there appears to be a consensus around the fact that maximal HR is decreased during FOR, NFOR and OTS. This is confirmed by our meta-analysis, as maximal HR was the single measure to be altered after both short-term (p = 0.01) and long-term (p = 0.007) increases in training load, thus emphasising its potential usefulness as a sign of FOR, NFOR and OTS.

### Post-exercise heart rate recovery

Post-exercise HR recovery has long been used as an objective sign of FOR, NFOR and OTS by clinicians. It is therefore not

---

**Table 1** Overall effect of overload intervention on heart rate and heart rate variability

<table>
<thead>
<tr>
<th>Categories</th>
<th>n</th>
<th>Overall SMD (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR Rest</td>
<td>244</td>
<td>0.18 (−0.02 to 0.38)</td>
<td>0.07</td>
</tr>
<tr>
<td>Submaximal exercise</td>
<td>193</td>
<td>−0.27 (−0.50 to −0.03)</td>
<td>0.03</td>
</tr>
<tr>
<td>Maximal exercise</td>
<td>189</td>
<td>−0.38 (−0.58 to −0.17)</td>
<td>0.0003</td>
</tr>
<tr>
<td>HRV LH+HF</td>
<td>83</td>
<td>−0.32 (−0.67 to 0.04)</td>
<td>0.08</td>
</tr>
<tr>
<td>HRV LF/HF</td>
<td>76</td>
<td>0.41 (0.08 to 0.73)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

HR, heart rate; HRV, heart rate variability; HF, high frequency; LF, low frequency; SMD, standardised mean difference.

---

**Table 2** Effect of duration of overload intervention on heart rate and heart rate variability

<table>
<thead>
<tr>
<th>Categories</th>
<th>≤2 weeks</th>
<th>≥2 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>SMD (95% CI)</td>
</tr>
<tr>
<td>HR Rest</td>
<td>89</td>
<td>0.55 (−0.09 to 1.02)</td>
</tr>
<tr>
<td>Submaximal exercise</td>
<td>61</td>
<td>−0.03 (−0.48 to 0.43)</td>
</tr>
<tr>
<td>Maximal exercise</td>
<td>26</td>
<td>−0.75 (−1.32 to −0.17)</td>
</tr>
<tr>
<td>HRV LH+HF</td>
<td>51</td>
<td>−0.49 (−1.03 to −0.04)</td>
</tr>
<tr>
<td>HRV LF/HF</td>
<td>51</td>
<td>0.52 (0.12 to 0.93)</td>
</tr>
</tbody>
</table>

HR, heart rate; HRV, heart rate variability; HF, high frequency; LF, low frequency; SMD, standardised mean difference.
unexpected to find it in the list of potential markers in numerous narrative studies.18–20 Surprisingly, we were not able to find published data that could be included in our meta-analysis. To our knowledge, the study by Verma et al.21 was the only one to provide indirect data supporting a delayed post-exercise recovery, through the measure of blood pressure. Urhausen et al.22 mentioned that post-exercise HR recovery was faster after deliberate prolonged overload training (p<0.05), but did not provide any data. With the exception of these opposing reports, all references cited in narrative reviews are either anecdotal observations or other narrative reviews. As long as it is not supported by experimental data, any conclusion about the validity of post-exercise HR recovery as a marker of FOR, NFOR and OTS will be hazardous. In this respect, it is worth noting that more recent reviews do not even mention its potential usefulness.3,6,11,12

Heart rate variability
HRV is a non-invasive measure of cardiovascular autonomic regulation.6 It is generally used in the medical literature as an independent predictor of all-cause mortality.6 Its use in the monitoring of FOR, NFOR and OTS is more recent, and relies on the distinction made by Israel64 between a sympathetic and a parasympathetic type of overtraining, according to the effects such overtraining has on the autonomic nervous system (ANS). Although there is little empirical or experimental evidence to support this classification, several studies have reported alterations in catecholamine concentrations at rest or during exercise after marked increases in training load.5–7 which could justify the potential usefulness of HRV in monitoring FOR, NFOR and OTS. In this meta-analysis, we found a small increase in the cardiovascular autonomic balance (LF/HF; p = 0.01), and a small but not significant decrease in total variability (LF+HF; p = 0.08). When considering the duration of interventions, this ANS alteration was limited to short-term overload, as it was followed by a moderate increase in LF/HF (p = 0.02) and a moderate but non-significant decrease in LF+HF (p = 0.07), whereas there were no modifications after interventions of >2 weeks (0.53 < p < 0.76). As for resting HR, cardiovascular autonomic balance as measured by HRV may be considered as a valid sign of short-term but not long-term fatigue.

Practical considerations
According to our results, resting HR, HRV and maximal HR may be used as markers of short-term fatigue (possibly FOR), whereas long-term fatigue (possibly NFOR/OTS) is expected to decrease HR during both submaximal and maximal workloads. From a practical point of view, it is important to determine if the effect of fatigue on these markers, which is primarily statistical in the context of a meta-analysis, is of sufficient magnitude to provide athletes, coaches and physicians with threshold values beyond (or under) which FOR, NFOR and OTS may be suspected. When using the standardised mean difference as a criterion, it appeared that these alterations were either small or moderate, but never large. To make sense of and convert these categories into HR data, we computed the weighted mean difference. As shown in table 3, in 100 athletes experiencing a performance decrease after a short-term overload training, 98 of them will have a resting tachycardia of 2–7 beats/minute (p = 0.0001). Considering that the day-to-day variability of resting HR has been reported to be ~5 beats/minute in healthy subjects with a mean value of 65 beats/minute,49 and that it can be much higher when the conditions are not strictly controlled,50 this means that overload-induced tachycardia will not be detectable in some cases. This difficulty in delineating between fatigue-induced alterations and day-to-day variability also applies to maximal HR. Bosquet et al.51 measured test–retest reliability of maximal HR in 50 endurance athletes and reported a standard error of measurement of 4 beats/minute. The overall 95% CI for the weighted mean difference in this meta-analysis was 2–6 beats/minute (table 3, p < 0.00001), suggesting that although statistically valid, maximal HR has few practical value for athletes, coaches or physicians in the prevention of FOR, NFOR and OTS. It is worth noting that we found a wider 95% CI after short-term overload training (3–12 beats/minute, p = 0.002), but it was computed from only 26 subjects, thus limiting its external validity. The same conclusion applies to HR during submaximal intensity exercise. The weighted mean difference we noted after long-term overload training (1–6 beats/minute, p = 0.002) was in the range of day-to-day variability of this measure. In fact, Lambert et al.52 reported a 95% CI of 6–8 beats/minute in 50 moderately trained subjects performing the same submaximal intensity test on 5 consecutive days.

According to our results, HRV is a potential marker of short-term fatigue, as we found a moderate increase in LF/HF after an overload training period of ≤2 weeks (table 2, p = 0.01). It was not possible to compute a weighted mean difference for this index because, depending on the algorithm used to compute it, mean LF/HF may vary from 0.52 (SD 0.70)44 to 153 (SD 82).46 If we consider that HRV measured in the frequency domain is moderately reliable,50 particularly LF/HF,51 it is obvious that the detection of such an effect requires a highly standardised protocol, including the control of training load in the preceding days,52 breathing frequency,51 time of day,52 digestion53 and all other stimuli that may increase the sympathetic nervous activity, such as temperature, luminosity, noise and the consumption of caffeine or alcohol. Efforts should also be made to fulfil the recommendations of the Task Force54 for the processing of interval data, including a sampling rate of >250 Hz for the detection of R waves, the replacement of ectopic beats, arrhythmic events, missing data and noise by interpolation from adjacent successive normal RR intervals, and the use of regularly sampled interpolation of the discrete event series for both parametric and non-parametric methods. In the absence of an objective criterion to interpret HRV alterations accurately during the follow-up of competitive athletes, any increase in LF/HF should be associated with other possible signs and symptoms of FOR, NFOR and OTS before diagnosing the

<table>
<thead>
<tr>
<th>Categories</th>
<th>n</th>
<th>WMD (95% CI), beats/minute</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 weeks</td>
<td>89</td>
<td>4.49 (–2.22 to 6.75)</td>
<td>0.0001</td>
</tr>
<tr>
<td>&gt;2 weeks</td>
<td>155</td>
<td>0.27 (–1.16 to 1.96)</td>
<td>0.62</td>
</tr>
<tr>
<td>Overall effect</td>
<td>244</td>
<td>1.99 (0.58 to 3.41)</td>
<td>0.006</td>
</tr>
<tr>
<td>Submaximal exercise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 weeks</td>
<td>61</td>
<td>–1.25 (–4.87 to 1.58)</td>
<td>0.39</td>
</tr>
<tr>
<td>&gt;2 weeks</td>
<td>132</td>
<td>–3.59 (–5.87 to –1.31)</td>
<td>0.002</td>
</tr>
<tr>
<td>Overall effect</td>
<td>193</td>
<td>–2.61 (–4.39 to –0.84)</td>
<td>0.004</td>
</tr>
<tr>
<td>Maximal exercise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 weeks</td>
<td>26</td>
<td>–7.47 (–12.21 to –2.74)</td>
<td>0.002</td>
</tr>
<tr>
<td>&gt;2 weeks</td>
<td>163</td>
<td>–3.62 (–5.60 to –1.64)</td>
<td>0.0003</td>
</tr>
<tr>
<td>Overall effect</td>
<td>189</td>
<td>–4.19 (–6.02 to –2.36)</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>
presence of a short-term fatigue (possibly FOR in this context) and adapting the training load to allow recovery.

Limitations
Correct interpretation of our data relies on the hypothesis that all participants included in this meta-analysis had at least FOR. Some studies provided diagnostic criteria and performance measures to objectively evaluate the level of fatigue of their participants, but others did not. An inclusion criterion could have been to select studies using this approach, but this would inevitably have deprived us of valuable information. Even if we assume this hypothesis was true, it remains to be determined as to what extent the fatigue induced by a relatively short period of overload training is comparable with fatigue occurring after years of high training load and psychological pressure. It is therefore acknowledged that the experimental model used to study the aetiology of FOR, NFOR and OTS is not optimum in this context and may have led to some shortcomings.

The aim of a meta-analysis is to combine the body of the literature on a given topic in order to provide evidence-based conclusions. In this study we identified only one additional moderator: the duration of the overload intervention (<2 weeks, >2 weeks). Other moderators could have been used, such as the strategy of increasing the training load (ie, intensity or volume) and the sex of the participants. A prerequisite to make valid comparisons is to have roughly the same number of subjects per moderator variables. This was clearly not the case for sex and overload strategy. Other measures could also have been included, particularly the use of orthostatic stress! to study more thoroughly fatigue-induced alterations of the cardiac autonomic control.24 27 43 Once again, the total number of subjects (25 subjects in this case, with different methods used to induce orthostatic stress) was not sufficient to address other moderator variables (ie, duration of the overload intervention) and thus to justify the inclusion of this measure in the context of a meta-analysis. Finally, this statistical method does not allow the use of case histories, whereas valuable reports have been published relating to FOR, NFOR and OTS.5 – 26 and the specific study of overtraining requires participants to be their own controls, thus excluding very interesting cross-sectional studies comparing athletes suffering FOR, NFOR and OTS with control subjects.26

CONCLUSION
The purpose of this meta-analysis was to test the effect of increased training load on HR or HRV, to determine whether they could be used as valid markers of FOR, NFOR and OTS. Although we found significant modifications after short-term fatigue (resting HR and LF/HF), long-term fatigue (HR during a submaximal workload) or both (maximal HR), the moderate amplitude of those alterations limits their clinical usefulness, as expected differences may fall within the day-to-day variability. Consequently, the correct interpretation of HR or HRV fluctuations during the training process requires the comparison of these markers with other objective signs and symptoms of FOR, NFOR and OTS.

Competing interests: None.

REFERENCES


CORRECTIONS

doi:10.1136/bjsm.2008.042200corr1


doi:10.1136/bjsm.2007.040915corr1