Physical Load and Referees’ Decision-Making in Sports Games: A Scoping Review

Nicolas Bloß, Jörg Schorer, Florian Loffing and Dirk Büsch
Institute of Sport Science, Carl von Ossietzky University Oldenburg, Oldenburg, Germany

Abstract
Referees in sports games have a high level of responsibility as they have to make correct and appropriate decisions at any point during a match. Regarding referees’ decision-making (RDM) as a perceptual-cognitive process, evidence suggests that physical load might reduce cognitive performance and thus might reduce RDM performance as well. In consideration of increasing game dynamics, referees have to cope with high physical load, but they have to make correct and appropriate decisions further on. Here, we review the current state of research on the relationship between physical load and RDM. A scoping review was performed, in accordance with the PRISMA guidelines, using the following databases: Scopus, PubMed, Web of Science (all databases), SURF and SPONET (both with an English and German search). Only primary studies written in English or German that investigated the relationship between physical load and RDM in sports games were included. Eleven studies included in the review investigated six physical parameters and RDM: match period, velocity, blood lactate, running time, heart rate, distance covered. Most findings of the studies showed no relationship between physical load and RDM \( (n = 18) \). Thirteen findings suggest a negative relationship and three findings indicate a positive relationship between physical load and RDM. Results of the scoping review show contradictory evidence across and within investigated different physical parameters. As RDM consists of multi-factorial components, it is recommended to conduct systematic research programs – field as well as experimental studies – to resolve the missing control of potential confounding variables and to consider the difference of internal and external load.

Key words: Performance, accuracy, stress, exertion, umpire, officiating.

Introduction
Referees in sports games have the responsibility to ensure that teams and players always abide by the sport-specific rules and intervene when a team or a player infringed a rule (Reilly and Gregson, 2006; Rullang et al., 2017). Therefore, a central task in officiating sports games is to make decisions about technical, offensive and defensive infringements (Bar-Eli et al., 2011; Plessner et al., 2009). As referees’ decisions can ultimately impact a game’s outcome (Larkin et al., 2014; McMahon et al., 2015), it is indispensable that referees’ decision-making (RDM) is on a high level (Helsen et al., 2019; McMahon et al., 2015). This includes that referees are able to apply the sport-specific rules during the entire match; that is, they need to have a well-developed physical fitness in order to keep up with the match dynamics when officiating (Kittel et al., 2019; McMahon et al., 2015; Plessner and McMahon, 2013; Reilly and Gregson, 2006). Mascarenhas et al. (2005) emphasize the relevance of both RDM and physical fitness for appropriate officiating when identifying five cornerstones of referees’ performance: (1) personality and game management, (2) physical fitness, positioning and mechanics, (3) knowledge and application of the law, (4) contextual judgement and (5) psychological characteristics of excellence. However, albeit highlighting the important role of both RDM and physical fitness, the authors’ model does not take into account potential relationships between cornerstones, RDM and physical fitness in particular whereby the central parameter of RDM is the correctness of the decisions made by the referees (Schweizer et al., 2011).

From this point of view, the relationship between RDM and physical load is of particular interest. Referees’ decisions are a perceptual-cognitive process (Helsen et al., 2019; Plessner and Haar, 2006; Schmidt et al., 2019), which might be impaired through high physical stress (Chang et al., 2012; Schmidt et al., 2019; Tomporowski, 2003). High physical stress can, for example, lead to an excessive enrichment of nor-epinephrine, which can result in a reduction of neuronal activity in the prefrontal cortex and ultimately cause a reduction in attention and executive functions (Arnsten, 2009; Schmidt et al., 2019). Thus, physical load may impair referees’ attentional control and, as a consequence, their decisions (Helsen et al., 2019; Schmidt et al., 2019).

Since decision-making is considered as an important task in officiating (Kittel et al., 2019), it is necessary to understand the relationship between physical load and RDM. Here we review the evidence on the relationship between physical load and RDM in sports games.

Methods
The scoping review was performed according to the Preferred Reporting Items for Systematic Reviews guideline (PRISMA; Moher et al., 2009).

Information sources and search strategy
The systematic search was carried out on September 4, 2019, by using the following databases: Web of Science (all databases), Scopus, PubMed, SURF and SPONET. Searches in SURF and SPONET were conducted in English as well as in German. The option ‘advanced searches’ was chosen for searches in Web of Science, PubMed and Scopus with regard to the articles’ title, abstract and keywords. “All Field” searches were executed in SURF and SPONET. The development and selection of the final keywords used was a process consisting of several search...
attempts and corresponding adjustments, which were dis-
cussed with all authors to improve the quality of the key-
words and the search. In the end, the following search
terms are used: "(referee* OR umpir* OR officiating) 
AND ("decision-making" OR performance OR judg*) 
AND (stress OR strain OR effort OR load OR exertion OR 
pressure)". The general German search term was: 
"(Schiedsrichte*) AND (Entscheid* OR Leist* OR Bew-
wert*) AND (Stress OR Belastung OR Anstrengung OR 
Beanspruchung OR Erschöpfung OR Druck)".

Eligibility criteria and study selection
To be included in the review, articles must strictly fulfil 
the following inclusion criteria. Articles have to be primary 
studies and investigate the relationship between physical 
load and RDM. They must also refer to sports games and 
be written in English or German. We set no limitation to 
the year of publication. Besides, articles were excluded 
when meeting defined exclusion criteria: articles are re-
views, abstracts, project descriptions, conference papers, 
interviews, theoretical papers or dissertations. Studies re-
ferring to athletes, studies without reference to sport games 
or articles not written in English or German were also ex-
cluded.

In accordance with the PRISMA approach (Moher et al., 2009), additional articles were appended in the 
selection process after the database search was conducted. 
Afterwards, duplicates were excluded and titles, abstracts 
and the written languages were screened and excluded 
when not meeting inclusion criteria. To ensure that no rel-
vant article was excluded, critical articles were kept up 
and were discussed by the authors. Finally, full-texts were 
analysed and discussed before exclusion. The remaining 
articles were integrated into the scoping review (see 
PRISMA flowchart in Figure 1).

Data extraction
For each study selected for inclusion in the review, the fol-
lowing parameters were extracted:
• Study characteristics: sport, sample, referee type, ex-
pertise (level of officiating, experience), sex, age and a 
brief report of the measurement
• Study results categorized by physical parameters and 
physical load (external or internal load), a short descrip-
tion of the results and its statistical key figures

As outcomes and reports of the studies were not uni-
form, data were processed for a scoping review. This type 
of review article aims at a first examination of the existing 
evidence on a research subject especially for elaborating 
future research (Tod, 2019).

Figure 1. Flow diagram (Moher et al., 2009).
**Risk of bias**
All studies included in this review were assessed with regard to risk of bias by the authors. To this end, we used the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018) to assess the methodological quality of included studies. Criteria and the assessment of the risk of bias of included studies are illustrated in Table 1.

With reference to the quality assessment with the MMAT, we follow the MMAT publishers’ recommendation to not report a numeric score to better inform about the quality of the studies and to avoid misunderstandings by discussing meanings of the ratings (Hong et al., 2018).

**Results**

**Study selection**
A total of \( n = 2,250 \) potential articles were identified through the literature search. Of those, \( n = 2,224 \) articles were found in the database search and \( n = 26 \) additional articles were included in the selection process, as required by PRISMA. Additional articles were obtained by screening references from relevant articles. After excluding

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Referee</th>
<th>Level</th>
<th>Experience</th>
<th>Sex</th>
<th>Age</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmed et al. (2017)</td>
<td>Futsal</td>
<td>18</td>
<td>R</td>
<td>Experts</td>
<td>&gt; 5y</td>
<td>Male</td>
<td>34.1±3.2y</td>
</tr>
<tr>
<td>Catteeuw et al. (2010)</td>
<td>Soccer</td>
<td>10</td>
<td>AR</td>
<td>Experts</td>
<td>FIFA: 2.6 ± 3.3y</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Elsworthy et al. (2014)</td>
<td>Australian football</td>
<td>29</td>
<td>Umpire</td>
<td>Experts</td>
<td>NR</td>
<td>NR</td>
<td>32.4±6.1y</td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>Rugby</td>
<td>8</td>
<td>R</td>
<td>Experts</td>
<td>7.3 ± 4.3y</td>
<td>NR</td>
<td>35.3±7.1y</td>
</tr>
<tr>
<td>Gomez-Carmona and Pino Ortega (2016)</td>
<td>Soccer</td>
<td>5</td>
<td>AR</td>
<td>Advanced</td>
<td>NR</td>
<td>NR</td>
<td>21.2±0.98y</td>
</tr>
<tr>
<td>Larkin et al. (2014)</td>
<td>Australian football</td>
<td>15</td>
<td>Umpire</td>
<td>Experts</td>
<td>235.2 ± 151.3 matches</td>
<td>NR</td>
<td>36±13.5y</td>
</tr>
<tr>
<td>Mallo et al. (2012)</td>
<td>Soccer</td>
<td>10</td>
<td>R</td>
<td>Experts</td>
<td>FIFA license: 9.2 ± 4.1y</td>
<td>39.4±2.4y</td>
<td></td>
</tr>
<tr>
<td>Mascarenhas et al. (2009)</td>
<td>Soccer</td>
<td>5</td>
<td>R</td>
<td>Experts</td>
<td>&gt; 4y</td>
<td>Male</td>
<td>38.2±5.89y</td>
</tr>
<tr>
<td>Oudejans et al. (2005)</td>
<td>Soccer</td>
<td>4</td>
<td>AR</td>
<td>Experts</td>
<td>&gt; 5y</td>
<td>Male</td>
<td>42.2y</td>
</tr>
<tr>
<td>Paradis et al. (2015)</td>
<td>Australian football</td>
<td>10</td>
<td>Umpire</td>
<td>Sub-elite</td>
<td>8.86 ± 2.04y</td>
<td>2.86±3.39y</td>
<td></td>
</tr>
<tr>
<td>Samuel et al. (2019)</td>
<td>Soccer</td>
<td>22</td>
<td>R</td>
<td>Experts</td>
<td>5-23y</td>
<td>Male</td>
<td>27.14±6.4y</td>
</tr>
</tbody>
</table>

R = referee, AR = assistant referee, EPL: English Premier League, y = years, m = matches, NR = not reported, ± = standard deviation
duplicates, \( n = 1,854 \) articles remained for further analyses. Subsequently, these articles were screened with reference to their written language, title and abstract. As a result, \( n = 1,842 \) articles were excluded, because they did not fulfill the inclusion criteria. The remaining \( n = 12 \) articles underwent a full-text analysis against the eligibility criteria. This led to further exclusion of \( n = 1 \) article. Finally, \( n = 11 \) articles were included in the qualitative analysis (see Figure 1).

### Sports games and referee type

Most of the included studies were carried out in soccer \((n = 6)\) followed by Australian football \((n = 3)\), rugby \((n = 1)\) and futsal \((n = 1)\). The participants’ referee type comprised umpires (Australian football), referees (soccer, futsal and rugby) and assistant referees (soccer). Participants officiated at least on an advanced level \((n = 24)\), but most participants were experts \((n = 179)\). Detailed study characteristics are listed in Table 1.

### Measurement

Included studies differ in the selected approach of analysing the relationship between physical load and RDM. In eight studies, video-analysis of recorded matches were conducted and thus RDM was analysed on-field (Ahmed et al., 2017; Catteuew et al., 2010; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mallo et al., 2012; Mascarenhas et al., 2009; Oudejans et al., 2005). With the exception of the study by Catteuew et al. (2010), selected video-sequences were assessed by experts (Ahmed et al., 2017; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mallo et al., 2012; Mascarenhas et al., 2009; Oudejans et al., 2005).

Three studies used video-based decision tests conducted in the laboratory (Paradis et al., 2015; Samuel et al., 2019) or in the half times of real matches (Larkin et al., 2014). The video-sequences were assessed by experts too (Larkin et al., 2014; Paradis et al., 2015; Samuel et al., 2019).

### Quality assessment

Four included studies were categorized as quantitative non-randomized studies (Larkin et al., 2014; Mascarenhas et al., 2009; Paradis et al., 2015; Samuel et al., 2019) and seven studies as quantitative descriptive studies (Ahmed et al., 2017; Catteuew et al., 2010; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mallo et al., 2012; Oudejans et al., 2005). Regarding the four quantitative non-randomized studies, two studies (Larkin et al., 2014; Mascarenhas et al., 2009) fulfil four of five criteria while the other two studies met three criteria (Paradis et al., 2015; Samuel et al., 2019). We could not assess criterion number 5 of the quantitative non-randomized studies (if the studies are administered as intended), as it was not explicitly reported. Assessing the other seven studies, all studies met three of five criteria (Ahmed et al., 2017; Catteuew et al., 2010; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mallo et al., 2012; Oudejans et al., 2005). Meanings of and reasons for this assessment will be reconsidered at a later stage of this review. An overview on the assessment of the risk of bias is given in Table 2.

### Aspects of physical load

Results of the included studies were summarized in Table 3. Eight studies investigated the relationship between match period and RDM (Ahmed et al., 2017; Catteuew et al., 2010; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Larkin et al., 2014; Mallo et al., 2012; Mascarenhas et al., 2009), followed by six studies that examined the relationship between referees’ velocity and RDM (EL; Catteuew et al., 2010; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mascarenhas et al., 2009; Oudejans et al., 2005). Three studies investigated the relationship between the referees’ heart rate and RDM (Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mascarenhas et al., 2009) and two studies assessed the relationship between the total distance covered and RDM (Emmonds et al., 2015; Oudejans et al., 2005).

### Table 2. Assessment of the risk of bias (appraised with the MMAT).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the sampling strategy relevant to address the research question?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Is the sample representative of the target population?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Are the measurements appropriate?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Is the risk of non-response bias low?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Is the statistical analysis appropriate to answer the research question?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Are the participants representative of the target population?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Are measurements appropriate regarding both the outcome and intervention (or exposure)?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Are there complete outcome data?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Are the confounders accounted for in the design and analysis?</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>During the study period, is the intervention administered (or exposure occurred) as intended?</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
</tr>
</tbody>
</table>

*Legend: 1 = yes, 0 = no; c = can’t tell*
Table 3. Results of the included studies.

<table>
<thead>
<tr>
<th>Load</th>
<th>Study</th>
<th>Description</th>
<th>p</th>
<th>d</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ahmed et al. (2017)</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; half vs. 2&lt;sup&gt;nd&lt;/sup&gt; half: 91.1 ± 14.9% vs 73.3 ± 17.4% correct decisions 1.7 ± 5.3% vs. 8.8% contentious decisions 7.2 ± 14.4% vs. 17.9 ± 17.1% incorrect decisions 24.8 ± 18.4% vs. 75.2 ± 18.4% missed fouls</td>
<td>0.007</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Categorical Distribution of errors over six 15-minutes periods</td>
<td>69% mistakes in 1&lt;sup&gt;st&lt;/sup&gt; half vs. 31% in 2&lt;sup&gt;nd&lt;/sup&gt; half</td>
<td>0.91</td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>Moderate increase of accuracy from minutes 40-50 to 50-60</td>
<td>NR</td>
<td>0.7</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Gomez-Carmona and Pino-Ortega (2016)</td>
<td>No correlation between heart rate (mean) and decision-making performance</td>
<td>NR</td>
<td>0.568</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Larkin et al. (2014)</td>
<td>Improvement of decision-making performance for quarter two to quarter four</td>
<td>NR</td>
<td>0.579</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Mallo (2012)</td>
<td>R: no uniform distribution of errors for six 15-min periods. Higher error percentage in last 15-min: 1&lt;sup&gt;st&lt;/sup&gt; 15-min= 7%</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; 15-min= 9%</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; 15-min= 12%</td>
<td>4&lt;sup&gt;th&lt;/sup&gt; 15 min= 17%</td>
<td>5&lt;sup&gt;th&lt;/sup&gt; 15-min= 9%</td>
</tr>
<tr>
<td>R: no uniform distribution of errors for six 15-min periods. Higher error percentage in last 15-min: 1&lt;sup&gt;st&lt;/sup&gt; 15-min= 6%</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; 15-min= 10%</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; 15-min= 10%</td>
<td>4&lt;sup&gt;th&lt;/sup&gt; 15 min= 19%</td>
<td>5&lt;sup&gt;th&lt;/sup&gt; 15-min= 25%</td>
<td>6&lt;sup&gt;th&lt;/sup&gt; 15-min= 21%</td>
</tr>
<tr>
<td>Samuel et al. (2019)</td>
<td>Lower accuracy in the second quarter compared to the third quarter</td>
<td>NR</td>
<td>0.36</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Samuel et al. (2019)</td>
<td>Lower accuracy in the fourth quarter compared to the third quarter</td>
<td>NR</td>
<td>0.24</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Mallo et al. (2012)</td>
<td>Less accurate in the first 15-min in each half, but not statistically significant: 1&lt;sup&gt;st&lt;/sup&gt; 15-min= 51%</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; 15-min= 69%</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; 15-min= 70%</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Mallo et al. (2012)</td>
<td>Uniform distribution of correct and incorrect decisions over six 15-minutes</td>
<td>NR</td>
<td>0.018</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Mallo et al. (2012)</td>
<td>No relationship between correct and incorrect decisions at instantaneous velocity at the time of the decision</td>
<td>NR</td>
<td>0.007</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Mallo et al. (2012)</td>
<td>Lower speed at incorrect decisions over the prior 5 seconds compared to correct decisions</td>
<td>NR</td>
<td>0.078</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>No relationship between referees’ velocity and decision-making accuracy</td>
<td>NR</td>
<td>0.36</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Gomez-Carmona and Pino-Ortega (2016)</td>
<td>Most errors at slow velocity of the R (35.2%): &gt;18 km/h ~ 4%</td>
<td>13-18 km/h ~ 8%</td>
<td>7.2-13 km/h ~ 19%</td>
<td>3.6-7.2 km/h ~ 29%</td>
<td>NR</td>
</tr>
<tr>
<td>Mascarenhas et al. (2009)</td>
<td>No relationship between referees velocity and decision-making accuracy</td>
<td>NR</td>
<td>0.02</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Mascarenhas et al. (2009)</td>
<td>No relationship between referees velocity and decision-making accuracy</td>
<td>NR</td>
<td>0.24</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Oudejans et al. (2005)</td>
<td>R make more errors when running or sprinting (&gt;8 km/h) compared when jogging or walking</td>
<td>NR</td>
<td>0.05</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Paradis et al. (2015)</td>
<td>No relationship between running time and decision-making performance</td>
<td>NR</td>
<td>0.05</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>No correlation between distance covered and decision-making accuracy</td>
<td>NR</td>
<td>0.02</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>No correlation between heart rate (mean) and decision-making accuracy</td>
<td>NR</td>
<td>0.2</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>No difference between heart rate (mean) and decision-making accuracy</td>
<td>NR</td>
<td>0.2</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Emmonds et al. (2015)</td>
<td>No difference between heart rate at correct (165.5 ± 12.5 bpm) and incorrect decisions (165.6 ± 13.3 bpm)</td>
<td>NR</td>
<td>0.1</td>
<td>NR</td>
<td></td>
</tr>
</tbody>
</table>

*Load: A = match period, B = velocity, C = running time, D = distance covered, E = heart rate, F = blood lactate.
†Abbreviations: R = referees, AR = assistant referees, AFO = Australian field officials, NR= not reported, p = statistical significance, d = effect size, r = correlation, ± = standard deviation

Finally, one study examined the relationship between blood lactate (Larkin et al., 2014) and RDM (Paradis et al., 2015).

**Match period.** Five studies found a negative relationship between match period and RDM, which means that with increasing playing time RDM decreased (Ahmed et al., 2017; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mallo et al., 2012; Samuel et al., 2019). Another five studies did not find evidence of a negative or positive relationship between match period and RDM (Ahmed et al., 2017; Catteeuw et al., 2010; Emmonds et al., 2015; Mascarenhas et al., 2009; Samuel et al., 2019).
Two studies revealed an increase of RDM with increasing playing time (Emmonds et al., 2015; Larkin et al., 2014).

Samuel et al. (2019; soccer) found a drop in RDM from the second to the third quarter (effect size: $d = 0.91$) as well as from the third to the fourth quarter ($d = 0.66$). Two studies, carried out in futsal (Ahmed et al., 2017) and soccer (Mallo et al., 2012), showed that RDM dropped in the second half of a game. In contrast, Samuel et al. (2019) did not find a difference between half times. Likewise, Ahmed et al. (2017) did not find a drop in RDM at contentions either. Three studies did not reveal a relationship between 10-minute or 15-minute periods and RDM (soccer: Catteeuw et al., 2010; rugby: Emmonds et al., 2015; soccer: Mascarenhas et al., 2009). However, two studies found a decrease in RDM in the last 10 minutes (rugby: $d = 0.86$, Emmonds et al., 2015) respectively 15 minutes (soccer: Mallo et al., 2012) of a match. At last, two studies revealed a positive relationship between playing time and RDM (rugby: Emmonds et al., 2015; Australian football: Larkin et al., 2014).

**Velocity.** The Relationship between referees’ velocity and RDM was analysed with the instantaneous referees’ velocity at the time of the decision (Catteeuw et al., 2010; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; soccer: $d = 0.02$, Mascarenhas et al., 2009; Oudejans et al., 2005) and with the running speed prior to the decision (Elsworthy et al., 2014a).

Examining first the relationship between the referees’ instantaneous velocity and RDM, two studies revealed a decrease of RDM with increasing velocity (soccer: Gomez-Carmona and Pino-Ortega, 2016; soccer: Oudejans et al., 2005) while four studies did not indicate a positive or negative relationship between the referees’ velocity and RDM (soccer: Catteeuw et al., 2010; Australian football: Elsworthy et al., 2014a; rugby: Emmonds et al., 2015; soccer: $d = 0.02$, Mascarenhas et al., 2009). A closer look at the two studies that reveal a negative relationship between the referees’ velocity and RDM indicates contradictory findings: Gomez-Carmona and Pino-Ortega (2016) found that most errors occur at slow velocity (<8 km/h; soccer), while Oudejans et al. (2005) found that errors occur more often at higher velocities (>8 km/h; soccer). As mentioned, four studies show no change of RDM in terms of the distribution of errors according to six different categories of speed (standing: 0.0 – 0.1 m/s, walking: 0.2 – 1.9 m/s, jogging: 2.0 – 3.9 m/s, running: 4.0 – 5.4 m/s, high-intensity running: 5.5 – 6.9 m/s, sprinting: >7.0 m/s; soccer: Catteeuw et al., 2010) nor to the instantaneous velocity at the time of the decision (Australian football: Elsworthy et al., 2014a; rugby: Emmonds et al., 2015; soccer: Mascarenhas et al., 2009).

At least, Elsworthy et al. (2014a) investigated the relationship between RDM and the referees’ velocity five seconds, thirty seconds, one minute and five minutes prior to the decision. Results indicate that the referees’ velocity five seconds prior to incorrect decisions was significantly higher compared to correct decisions.

**Running time.** Paradis et al. (2015) investigated the relationship between running time and RDM in Australian football (Australian football: Paradis et al., 2015), but they did neither find a negative nor positive relationship.

**Distance covered.** No relationship was found between the total distance covered and RDM (rugby: Emmonds et al., 2015; soccer: $d = 0.27$, Mascarenhas et al., 2009).

**Physiological parameters.** Two studies did not find a relationship between heart rate and RDM (rugby: Emmonds et al., 2015; soccer: $d = 0.01$, Mascarenhas et al., 2009), while one study revealed a drop in RDM especially when the heart rate was above 85% (soccer: Gomez-Carmona and Pino-Ortega, 2016). One study examined physical exertion with blood lactate did not find evidence for a relationship between physical exertion and RDM (Australian football: Larkin et al., 2014).

Overall, most findings in this review did not reveal a negative or positive relationship between physical load and RDM ($n = 18$; Ahmed et al., 2017; Catteeuw et al., 2010; Elsworthy et al., 2014a; Emmonds et al., 2015; Larkin et al., 2014; Mascarenhas et al., 2009; Paradis et al., 2015; Samuel et al., 2019). However, $n = 13$ findings point to a negative relationship between physical load and RDM (Ahmed et al., 2017; Elsworthy et al., 2014a; Emmonds et al., 2015; Gomez-Carmona and Pino-Ortega, 2016; Mallo et al., 2012; Oudejans et al., 2005; Samuel et al., 2019), while only two studies suggest an increase of RDM with increasing load (match period; Emmonds et al., 2015; Larkin et al., 2014).

**Discussion**

**Empirical evidence**

Recent studies examined the physical demands imposed on referees across various sports games, especially in soccer (Castagna et al., 2007) and Australian football (Elsworthy et al., 2014b). Current trends in competitive sports show increasing game dynamics and thus altered physical load not only for players but also referees (Bilge, 2012; Dolański et al., 2017; Mascarenhas et al., 2005; Weston et al., 2007). Hence, referees are required to cope with physical load (Brightmore et al., 2016; Dolański et al., 2017; Leicht, 2008; Martin et al., 2001; Weston et al., 2007) when making decisions in a match or day by day in a tournament (Helsen and Buljtynck, 2004). Thus, it is important to focus on referees’ physical fitness and the relationship between physical load and RDM, as RDM might decrease with increasing physical load (Helsen et al., 2019; Schmidt et al., 2019). Therefore, the aim of this review was to provide an overview of studies examining the relationship between physical load and RDM. Our findings suggest that empirical evidence is unclear and contradictory.

**Methodological assessment**

With regard to the risk of bias, an advantage of the seven studies conducting video-analyses is that the referees made decisions in real-world situations. Thus, these studies probably have high external validity. However, such approach does not enable to control and vary the independent variables nor to identify and control confounding variables. Consequently, the above seven studies’ conclusions on the relationship between physical load and RDM have to be treated carefully as the results might be impaired due to
confounding variables, for example crowd noise (e.g., Unkelbach and Memmert, 2010). Lab-based studies, in turn, may ensure high internal validity by identifying and controlling confounding variables, systematically varying the independent variable(s) and minimizing error variance (Aziz, 2017). Hence, lab-based studies can provide first understanding of the suggested systematic relationship between physical load and RDM as well as about confounding variables. Overall, to comprehensively add to both theory and practice, and reduce the risk of bias, we recommend a combination of a lab-based and field-based. Both have specific advantages, in systematic research programs comprising several studies which build upon each other (Büsch, 2019). Research programs should base on theoretical models that allow deriving testable hypotheses.

Furthermore, most studies examined the relationship between external load and RDM rather than internal load and RDM. This may be problematic because external load is more a kind of description of a physical task (e.g., distance). Internal load, however, is an individual’s psychophysiological response to external load (Impellizzeri et al., 2019). Thus, different individuals (e.g., referees) can experience the same external load (e.g., operationalized by match period) as differently exhausting (Impellizzeri et al., 2019). Therefore, we recommend that future studies should test the hypothesized relationship between physical load and RDM with consideration of external and internal load. By controlling the internal load, it can be ensured that an individualized external load leads to a comparable internal load, e.g. relative heart rate (Impellizzeri et al., 2019), in order to examine the relationship between physical load and RDM inter-individually and to make intra-individual comparisons more reliable.

Limitations
The present review does not provide a clear picture of evidence regarding the relationship between physical load and RDM. First of all, this might be due to the low number of studies included in this review. Furthermore, studies were conducted in different sports and this may render comparability of results potentially difficult, although it also remains open whether there are any sport-specific differences at all. Likewise, consideration of different referee types (e.g., referee or assistant referee in soccer) across studies might have enhanced heterogeneity in findings, assuming that different referee types are faced with different physical as well as decision-making demands (Mallo et al., 2012).

Conclusion
Based on this scoping review, we conclude that the current state of evidence on the relationship between physical load and RDM is contradictory and thus inconclusive. This refers to evidence across and within physical parameters. Finally, in view of the risk of bias described above, we suggest a combination of lab- and field-based approaches as part of systematic research programs aiming to determine the impact of selected physical load parameters on RDM under naturalistic conditions.

Acknowledgements
This research was funded from the Federal Institute of Sport Science based on a resolution of the German Bundestag (official file number: ZMV14-071001/19/20). The authors have no conflict of interest directly relevant to the content of the study. The authors declare the present study comply with the current laws of the country in which they were performed.

References


Hong, Q.N., Fábregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B.,


Key points

- Officiating in sport games requires high-level decision-making under various constraints.
- A scoping review was conducted on the relationship between physical load and RDM.
- Most studies suggest that physical load is not or negatively correlated with RDM.
- Conclusions, however, are limited due to the low internal validity of the included studies.
- Systematic research programs with targeted isolated testing of the relationship between physical load and RDM are recommended.

AUTHOR BIOGRAPHY

Nicolas BLOß
Employment Institute of Sport Science, Carl von Ossietzky University Oldenburg
Degree PhD Student, M.A.
Research interests judgement and decision-making in sport, sport performance, cognition, exercise science, sports sciences
E-mail: nicolas.bloss1@uni-oldenburg.de

Jörg SCHORER
Employment Institute of Sport Science, Carl von Ossietzky University Oldenburg
Degree Full-professor, PhD
Research interests Talent in sport, strength and conditioning, expertise, cognition, perception, motor learning and motor control
E-mail: joerg.schorer@uni-oldenburg.de

Florian LOFFING
Employment Institute of Sport Science, Carl von Ossietzky University Oldenburg
Degree PhD
Research interests Laterality, expertise, cognition, perception, motor learning and motor control
E-mail: florian.loffing@uni-oldenburg.de
Dirk BÜSCH

Employment
Institute of Sport Science, Carl von Ossietzky University Oldenburg

Degree
Full-professor, PhD

Research interests
Exercise performance, exercise science, sports sciences, exercise testing, exercise physiology, sport physiology, sport biomechanics, muscle physiology, talent in sport

E-mail: dirk.buesch@uni-oldenburg.de

Nicolas Bloß

Institute of Sport Science, Carl von Ossietzky University of Oldenburg, Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany