Journal of Electromyography and Kinesiology 47 (2019) 65–87

Contents lists available at ScienceDirect

Journal of Electromyography and Kinesiology

journal homepage: www.elsevier.com/locate/jelekin

Review

Diagnostic accuracy, validity, and reliability of Tensiomyography to assess muscle function and exercise-induced fatigue in healthy participants. A systematic review with meta-analysis☆

C. Lohr[a,b,*], T. Schmidt[a,b,c,d], I. Medina-Porqueres[e], K.-M. Braumann[a], R. Reer[a], J. Porthun[f]

[a] Department of Sports Medicine, Faculty for Psychology and Human Movement Science, Institute for Human Movement Science, University of Hamburg, Hamburg, Germany
[b] Osteopathie Schule Deutschland (OSD), Hamburg, Germany
[c] Osteopathic Research Institute (ORI), Hamburg, Germany
[d] Department of General Practice, University Medical Center, Göttingen, Germany
[e] Department of Physical Therapy, Faculty of Health Sciences, University of Malaga, Malaga, Spain
[f] Department of Health Sciences, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Gjøvik, Norway

ARTICLE INFO

Keywords:
Mechanomyography
Muscle fatigue
Validity
Gender bias
Systematic review

ABSTRACT

Tensiomyography™ (TMG) is a non-invasive method to monitor skeletal muscle mechanical characteristics. This systematic review and meta-analysis reports on diagnostic accuracy, validity, and reliability of TMG (maximal radial displacement [Dm], contraction time [Tc], delay time [Td], and velocity of contraction [Vc]) to assess exercise-induced fatigue in healthy volunteers, with the specific aim to determine the current level of supporting evidence. Systematic literature searches within Medline, Embase and Sportdiscus databases were conducted from January 1990 through November 2018. Methodological quality was evaluated by the *Assessment of Diagnostic Accuracy Studies* (QUADAS) tool or *the Validity and Reliability Critical Appraisal Tool* (CAT) or the *Quality Appraisal of Diagnostic Reliability checklist* (QAREL). Meta-analytical methods were utilised to summarize relative reliabilities of Dm, Tc, Td (95%, CI). The methodological quality of the 19 included studies (n = 373; female = 13.0%) ranged from low to high quality. The analysis revealed insufficient diagnostic accuracy and validity, mixed results regarding absolute reliability, and high to excellent relative reliability for the assessed measures. To conclude, robust evidence for diagnostic accuracy/validity of TMG has yet to be determined, whereas there is substantial evidence for its reliability. Higher methodological standards need to be established, including the avoidance of gender bias.

1. Introduction

The non-invasive assessment of alterations in skeletal muscle function as a consequence of athletic performance or in a clinical setting has been substantially examined (Andonian et al., 2015; Goubert et al., 2017; Langevin et al., 2011; Liederbach et al., 2014; Lin et al., 2016). Here, one emphasis of complementary research lies on quantifying the effect of peripheral fatigue-related changes in electrophysiological and dynamical muscle characteristics (Cè et al., 2017; Kallenberg et al., 2007; Tosovic et al., 2016; Walker et al., 2012; Yoshitake et al., 2001). Muscle fatigue – defined as any decline in muscle capacity associated with muscle activity (Allen et al., 2008) – is complex and multifactorial. In the further course of this study the focus will be on exercise-induced muscle fatigue characterized as a reversible loss in maximal force or power production as a consequence of work over time (Finsterer, 2012; Gosker and Schols, 2008). Muscle fatigue is usually quantified by maximal voluntary contraction (MVC), which is considered the 'gold-standard' of fatigue (Place et al., 2007). To monitor acute effects of muscle fatigue on mechanical properties of muscle contraction without the assessment itself causing additional fatigue, surface mechanomyographic (MMG) methods have been applied (Orizio et al., 2003; Tosovic et al., 2016; Yoshitake et al., 2001). These non-invasive techniques are regarded as the mechanical counterpart of surface electromyography (EMG) (Gordon and Holbourn, 1948; Orizio et al., 2003). MMG records and quantifies lateral vibrations and thickening of muscle fibre at the surface of a muscle by specific sensors (transducers) during voluntary or

☆ AREA OF EXPERTISE: Evaluation of muscle contraction through mechanomyographic detection technique.

* Corresponding author at: Eppendorfer Landstraße 42, 20249 Hamburg, Germany.
E-mail address: praxis@christinelohr.de (C. Lohr).

https://doi.org/10.1016/j.jelekin.2019.05.005
Received 9 October 2018; Received in revised form 16 March 2019; Accepted 3 May 2019

evoked contraction (Beck et al., 2006; Orizio, 1993). Several transducers have been utilised to detect the surface MMG signal such as piezoelectric contact sensors and accelerometers (Cè et al., 2017; Watakabe et al., 1998), laser distance sensors (Than et al., 2018), and microphones (Yoshitake et al., 2005). Despite encouraging results that have been achieved by using MMG techniques, there are some challenges mostly of technical nature, for example a low signal-to-noise ratio and resulting high variability, cumbersome equipment, and necessary post-processing of the signals (Orizio, 2002; Wong, 2001).

An alternative and easy to handle MMG method was introduced named Tensiomyography™ (TMG) about twenty years ago (Dahmane et al., 2001; Kersevan et al., 2002; Valenčič and Knez, 1997). In contrast to other MMG methods, TMG only functions with stimulated contractions. The device incorporates a high-precision displacement sensor (4 μm) which is placed perpendicularly with a slight pre-tension (0.2 N/cm^2) onto the surface of the muscle belly in order to record the radial deformation of the muscle following electrically evoked stimulation. This pre-tension is considered to be one fundamental methodological difference between TMG and other MMG techniques (Simunic et al., 2010). In fact, as Krizaj et al. (2008) stated, this controlled pre-tension is the major benefit of TMG compared to other MMG methods, since it has a positive impact on the assessment of contraction dynamics. The electrical stimulation is delivered by two self-adhesive electrodes being placed symmetrically proximal and distal at a defined distance to the displacement sensor. In order to attain maximal individual response, the stimulation is progressively increased until the individual maximal radial displacement amplitude is reached (Simunic et al., 2011). In a direct comparison between two different transducers (contact vs. laser sensor), Seidl et al. (2017) found that both sensors were reliable for the assessed time-based parameters as well as maximal muscle displacement, with the contact sensor (which is similar to the one applied in TMG) recording higher mean values on average. Nevertheless, the authors considered this difference as clinically irrelevant.

Approximately 130 peer reviewed papers on TMG measurements have been released during the last decades. This non-invasive and easy to handle device is under discussion for a variety of applications. One focal view is set on various fields of physical activity (e.g. specific adaptation to exercise and/or age, determination of fibre type distribution, muscle fatigue, and recovery) (Dahmane et al., 2005; Paula Simola et al., 2016a,b; Peterson and Quiggle, 2016; Šimunič et al., 2018). Another focus, albeit to a lesser extent, lies on (neuro-)muscular disorders (Gasparini et al., 2012; Grabljevec et al., 2005; Neamtu et al., 2016; Rusu et al., 2016; Seijas et al., 2016). To guarantee comparability between the investigations, this systematic review will focus on studies related to sport scientific research performed in both male and female healthy subjects.

TMG obtains five primary measures; one spatial measure expressed in millimeters (mm), and four temporal measures expressed in milliseconds (ms), which can be derived from the twitch displacement-time curve: muscle displacement (Dm), delay time (Td), contraction time (Tc), half-relaxation time (Tr), and sustain Time (Ts). Within the framework of this study, the main emphasis will be on the most evaluated primary TMG parameters, with the highest level of reliability (Krizaj et al., 2008; Martín-Rodríguez et al., 2017a; Tous-Fajardo et al., 2010): the spatial variable Dm and the temporal variables Td, and Tc. Dm is considered to provide information about muscle fatigue, muscle stiffness and early atrophic processes (Ditroilo et al., 2011; Hunter et al., 2012; Pisot et al., 2008; Pisot et al., 2016). Td reflects the time from the onset of electrical stimulation until 10% of Dm, also characterized as reaction or activation time (Alentorn-Geli et al., 2015). Tc is defined as the time lapse between 10 and 90% of the peak value of Dm. This measure is commonly linked to muscle fibre type composition (Dahmane et al., 2001; Simunic et al., 2011), and speed of force generation (Loturco et al., 2016).

Alongside these primary measures, an indirect approach to assess the velocity of muscle radial deformation (Vc) has been reported increasingly (Gasparini et al., 2012; Lohr et al., 2018a; Loturco et al., 2016; Macgregor et al., 2016; Macgregor et al., 2018a; Paula Simola et al., 2015; Raeder et al., 2016; Wiewelhove et al., 2015; Wiewelhove et al., 2017). As several authors stated (Loturco et al., 2016; Paula Simola et al., 2015), alterations of contraction time might be influenced by the extent of radial muscle displacement, and as a consequence, hampers an appropriate interpretation of the results (Paula Simola et al., 2015). Vc is delineated as the rate of muscle contraction or rate of deformation development, and in fact, to date there is no standardized method of calculation for this secondary TMG raw data-derived parameter. Some authors integrate Dm, Tc, and Td within their calculation: mean velocity until 10%Dm (V_{10}), or mean velocity until 90% Dm (V_{90}) (Paula Simola et al., 2015; Paula Simola et al., 2016b; Raeder et al., 2016; Wiewelhove et al., 2017). Loturco et al. (2016) provide an alternative approach by dividing Dm through the sum of Td and Tc. A further procedure to assess the velocity of contraction is defined as the rate (mm/s) of contraction between 10 and 90% of maximal Dm (Vc = Dm80%/Tc) (Lohr et al., 2018a; Macgregor et al., 2016; Macgregor et al., 2018a). Thus the combination of several TMG-derived parameters to a single index potentially provides a more substantial evaluation of specific training interventions, for example peripheral muscle fatigue.

In the course of the implementation process of a novel diagnostic and treatment monitoring tool, it is vital to ensure that this tool has been successfully validated. Therefore the measurements must be accurate, valid, and reliable. Diagnostic accuracy describes the ability of an instrument or test to correctly discriminate cases with a certain condition and cases without the condition (Zhou et al., op. 2011). The validity of an instrument refers to the degree to which the interpretations of the results of a test are assured. This is dependent on the test's proposed uses (Goodwin and Leech, 2003). Reliability provides information on the stability, consistency and reproducibility of a measurement device or test (Kimberlin and Winterstein, 2008; Weir, 2005). Reliability is best estimated by more than one index, and therefore, it is recommended to combine indexes of relative (e.g. ICC) and absolute (e.g. SEM, CV) reliability (Atkinson and Nevill, 1998).

During the last decades, numerous studies have been published on TMG validity and reliability, whereas investigations on TMG diagnostic accuracy have only begun to be carried out recently. Up to the present two reviews on TMG have been released (Macgregor et al., 2018b; Martín-Rodríguez et al., 2017a). Martín-Rodríguez et al. (2017a) previously conducted a qualitative systematic review of studies examining the reliability and measurement error of TMG whereas Macgregor et al. (2018b) provided a critical and comprehensive overview of the existing evidence on tensiomyographic application as an assessment device within sport medical supervision. The present systematic review with meta-analysis expands upon these prior works in multiple ways: (1) We systematically summarized information on TMG studies of diagnostic accuracy and validity to determine exercise-induced muscle fatigue in healthy male and female volunteers with the TMG measures Dm, Tc, Td, and Vc. (2) We focused on the instrument's reliability for detecting spatial (Dm) and contractile characteristics (Tc, Td, and Vc), and furthermore, quantitatively assessed the data on relative reliability of three of the mentioned TMG measures (Dm, Tc, and Td). (3) We scrutinised the methodological quality and risk of bias of the studies included by using the *Quality Assessment of Diagnostic Accuracy Studies* (QUADAS) tool (Whiting et al., 2003), the V*alidity and Reliability Critical Appraisal Tool* (CAT) (Brink and Louw, 2012) or the *Quality Appraisal of Diagnostic Reliability* Checklist (QAREL) (Lucas et al., 2010), depending on the study type. (4) We summarized the overall level of evidence for the assessed TMG parameters in terms of diagnostic accuracy, validity, and reliability. There is hitherto no systematic review on the state of studies on accuracy and validity of Tensiomyography™ to observe exercise-induced muscle fatigue. The present investigation intends to fill this gap.

2. Methods

2.1. Registration

The study was registered with the International Prospective Register of Systematic Reviews (PROSPERO# CRD42018087126) (Lohr et al., 2018b).

2.2. Study design

This systematic review and meta-analysis was compiled applying recommendations on the conduct and reporting of systematic reviews and meta-analyses outlined in the *Preferred Reporting Items in Systematic Reviews and* Meta-*Analyses* (PRISMA) guidelines (Moher et al., 2010).

2.3. Search strategy

Using the interfaces Pubmed, OVID, and EBSCO a computer assisted systematic literature search of Medline, Embase, and Sportdiscus databases was performed by one author (CL) from 01 January through 04 April 2017. The rationale for selecting Medline and Embase is based on the fact that the combination of these two most important databases in sports medicine and orthopaedic surgery leads to a high detection rate of primary research (Slobogean et al., 2009). Additionally, it is recommended to search in a content-specific database (e.g. Sportdiscus) (Prinsen et al., 2018). The search was augmented using reference lists of relevant articles to assure that as many related papers as possible were captured. Grey literature was searched unsystematically. The literature search was updated on 30 November 2018. Search filters were developed using a combination of keywords and medical subject headings (MeSH)/Emtree. The selected terms were combined in Medline and Embase using the Boolean operator 'AND' or 'OR' A similar search strategy was carried out in the database Sportdiscus. In order to avoid an exclusion of relevant articles, methodological search filters were not included (Doust et al., 2005; Leeflang et al., 2013; Whiting et al., 2011). The search strategy is outlined in detail in Appendix A.

2.4. Screening and eligibility criteria

One reviewer (CL) performed the screening process in Microsoft Excel 2010, version 14.0, using predetermined inclusion and exclusion criteria. Articles were eligible if they contained the following criteria: (1) male and/or female healthy adults and/or adolescents were included; (2) the core subject of the investigation was the instruments' diagnostic accuracy regarding the ability to discriminate between fatigued and recovered subjects and/or the validity of TMG measures to assess exercise induced muscle fatigue and/or the relative and/or absolute reliability of TMG measures; (3) peer reviewed observational, cross-sectional, validity or controlled trial research studies were included; (4) the study was published in English.

Exclusion criteria were as follows: (1) the study focussed on participants with neurological or orthopaedic disorders; (2) testing was performed on infants; (3) diagnostic accuracy and/or relative and/or absolute reliability of TMG measures were not reported; (4) non-peer-reviewed studies, conference proceedings, editorials, books, theses and expert opinions; (5) non-human; (6) non-English.

A PRISMA flow chart was used to illustrate the study screening and selection process (Fig. 1).

2.5. Data extraction

Information and data extraction was carried out independently by one reviewer (CL) and verified by a second reviewer (JP) utilizing a predesigned extraction sheet. Demographical data regarding the study population characteristics and health status were identified as well as information concerning the testing circumstances. Then, diagnostic reference standard and statistical measures on diagnostic accuracy and/or validity, and/or reliability were extracted. A detailed listing and description of the extracted key parameters are shown in Table 1. Disagreements among the reviewers were resolved through discussion or by referral to a third reviewer (TS). Authors were contacted when necessary for additional information.

2.6. Study quality and risk of bias assessment

Each of the full text articles was assessed independently by two reviewers (CL, JP) assessing the study quality and risk of bias. For studies of diagnostic accuracy the *Quality Assessment of Diagnostic Accuracy Studies* QUADAS tool (Whiting et al., 2003) was utilized, and for validity studies the *Validity and Reliability Critical Appraisal Tool* (CAT) (Brink and Louw, 2012) was used, whereas the quality evaluation of reliability studies was operationalized through the *Quality Appraisal of Diagnostic Reliability* Checklist (QAREL) (Lucas et al., 2010).

The QUADAS tool encompasses 14 items, each of which having a 'yes', 'no' or 'unclear' answer option. Since the QUADAS evaluation tool does itself not incorporate a quality score (Whiting et al., 2005), several studies have used this tool and implemented a stratification of scoring low risk of bias and high quality if the QUADAS score was 10 or greater, whereas a score below 10 was associated with high risk of bias and low quality (e.g., Cook et al., 2007; Cook et al., 2012; Hegedus et al., 2007; Pacheco-Carrillo and Medina-Porqueres, 2016).

The QAREL checklist is especially conceptualized for the quality appraisal of diagnostic reliability studies (Lucas et al., 2010). This rating tool has been validated (Lucas et al., 2010; Lucas et al., 2013), and to date, QAREL has been applied in numerous systematic reviews (e.g., Boswell et al., 2015; Carlsson and Rasmussen-Barr, 2013; Falco et al., 2012; Lange et al., 2017). The checklist includes 11 items, which can be rated with 'yes', 'no', 'unclear' or 'not applicable', indicating individual ratings for each item, rather than elaborating an overall score. In an adaptation of Carlsson and Rasmussen-Barr (2013), we stipulated a stratification of scoring low risk of bias and high quality if > 72% of the applicable questions were rated 'yes'. A moderate risk of bias and moderate quality was considered with positive ratings above 63% and up to 72%, whereas a high risk of bias and low quality was associated with a positive score of 63% or below.

The CAT tool has been developed specifically to enhance the methodological quality of reporting of validity and reliability studies evaluating objective clinical tools (Brink and Louw, 2012). Five of a total of 13 items relate to both validity and reliability studies, whereas four items refer to validity studies only and four items to reliability studies only. Each item has a 'yes', 'no' or 'not applicable' answer option. In case of validity studies, the items which relate to reliability should be rated as 'not applicable', and vice versa (Brink and Louw, 2012). Although CAT has been developed to evaluate the methodological quality of both, studies on reliability and validity, we decided to utilize this appraisal tool only for the incorporated papers on validity. The reason for this is that QAREL has been used more frequently for the evaluation of reliability studies and therefore ensures a better comparability with other reviews. CAT does not provide an overall quality score, therefore we followed Prowse et al. (2016) and used their total methodological quality ratings for validity studies: 0–2 poor, 3–5 fair, 6 or 7 moderate, and 8 or 9 high. On this basis we specified a high potential risk of bias if the CAT score was 5 or below, a moderate risk of bias was considered with a score of 6 or 7, and a score of 8 or above was rated as a low potential risk of bias.

Disagreements among the reviewers regarding the results of the QAREL, QUADAS, CAT assessment were discussed or resolved by a third reviewer (TS).

2.7. Meta-analysis of relative reliability measures

Due to the heterogeneous designs of the included studies on

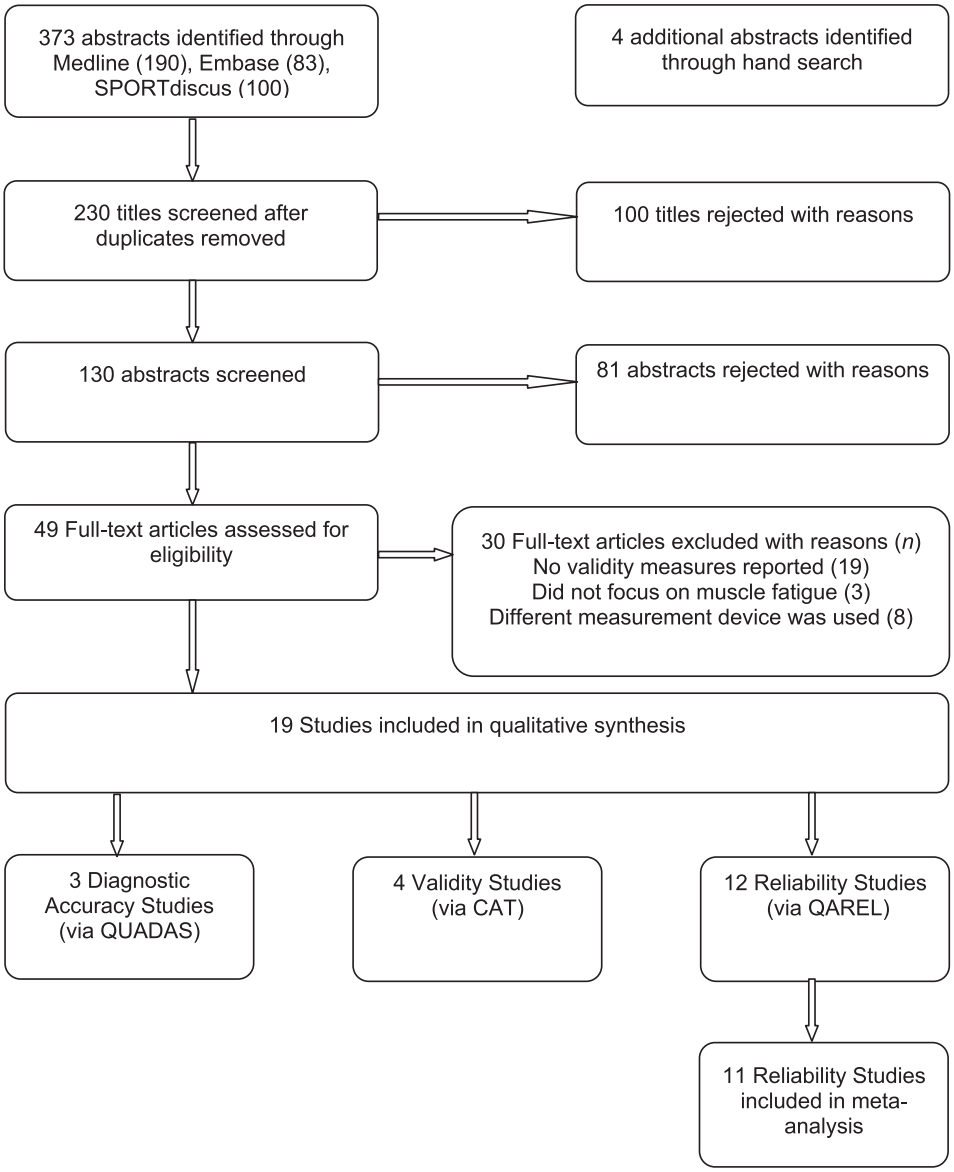

Fig. 1. Flow diagram of study selection. *Abbreviations:* CAT, validity and reliability Critical Appraisal Tool; QAREL, Quality Appraisal of Diagnostic Reliability checklist; QUADAS, Quality Assessment of Diagnostic Accuracy Studies Tool.

diagnostic accuracy and validity, a meta-analysis could only be performed for the reviewed reliability studies. The analysis was conducted using R version 3.4.0 with the aid of Meta package. Some studies report several values for reliability. These values (within a study) are based on the identical sample and thus, they are connected with each other. Random effects models were used to summarize the relative reliability indices (ICCs) and respective confidence intervals (CI 95%) (Borenstein et al., 2009). For this purpose a Fisher z-transformation was conducted for each individual score in order to calculate the arithmetic mean for every single study. Subsequently, these weighted average values were converted back into the original scale. Consequently, only one reliability value per study was taken into account in the meta-analysis. This proceeding ensures that studies reporting more than one value for reliability will have no stronger impact within the analysis. To evaluate the heterogeneity between the individual studies, tau-squared (τ^2) was conducted, particularly considering the high precision of the values being pooled (Rücker et al., 2008). Each study was weighted by the inverse variance method, and the heterogeneity was estimated by the method of Hunter and Schmidt (2006).

2.8. Quality assessment of accuracy, validity and absolute reliability measures

For the measures on diagnostic accuracy, validity, and absolute reliability we evaluated the quantitative results of each measurement property using quality criteria adapted from Terwee et al. (2007), Prinsen et al. (2018), and Mokkink et al. (2018). Each individual result of diagnostic accuracy, validity, and absolute reliability measures was rated using the criteria: sufficient (+), indeterminate (?), and insufficient (−) as outlined in Table 1.

2.9. Best evidence synthesis

Two reviewers (CL, JP) independently established the overall level of evidence by synthesizing the risk of bias scores (QUADAS, CAT, QAREL) of each individual study and the qualitative analysis of the measurement properties. For this purpose the following definitions based on van Tulder et al. (2003) were applied, see also Martín-Rodríguez et al. (2017a): strong (consistent findings in multiple studies of good methodological quality or in 1 study of excellent

Table 1
Definition and Interpretation of extracted measures.

Measurement property		Definition	Criteria* for a good measurement property and Quality rating	
Diagnostic Accuracy	[1]Sensitivity (SN)	Percentage of fatigued subjects who are correctly identified by TMG.	+	≥ 80%[2]
			−	< 80%[2]
			?	Not all information for '+' are reported
	[1]Specitivity (SP)	Percentage of non-fatigued subjects who are correctly identified by TMG.	+	≥ 80%[2]
			−	< 80%[2]
			?	Not all information for '+' are reported
	[1]Positive predictive value (PPV)	Percentage of subjects classified as fatigued by TMG who are in fact fatigued.	+	≥ 80%[2]
			−	< 80%[2]
			?	Not all information for '+' are reported
	[1]Negative predictive value (NPV)	Percentage of subjects classified as non-fatigued by TMG who are in fact non-fatigued.	+	≥ 80%[2]
			−	< 80%[2]
			?	Not all information for '+' are reported
	[1]Area under the receiver operating (ROC) curve (AUC)	The larger the area under the ROC the greater is the discriminative power of TMG (range from 0.5 to 1).	+	AUC ≥ 0.8
			−	AUC < 0.8
			?	Not all information for '+' are reported
	[3]Youden's index (*J*)	The ratio of anticipated TMG result in fatigued subjects to the non-fatigued subjects (0 = relatively poor diagnostic accuracy of a test, 1 = relatively large diagnostic accuracy of a test).	+	1
			−	0
			?	Not all information for '+' are reported
	[4]Diagnostic effectiveness (DE, %)	Percentage of all subjects who were correctly categorized by TMG.	+	≥ 80%
			−	< 80%
			?	Not all information for '+' are reported
	[4]Misclassification rate (MR, %)	Percentage of all subjects who were incorrectly categorized by TMG.	+	≤ 80%
			−	> 80%
			?	Not all information for '+' are reported
Validity	[5]Criterion validity	Refers to the extent to which TMG results relate to a gold-standard. [6]Concurrent validity: TMG and criterion measures are taken at the same time. [6]Predictive validity: the criterion measurement is performed at a later stage.	+	Correlation with gold standard ≥ 0.70
			−	Correlation with gold standard < 0.70, despite adequate design and method
			?	Doubtful design or method
	[5] Construct validity	Refers to the extent to which TMG results relate to other measures of fatigue in a manner that is consistent with theoretically deduced hypotheses concerning exercise-induced fatigued.	+	≥ 75% of the results are in accordance with the hypothesis
			−	< 75% of the hypothesis was not confirmed, despite adequate design and methods
			?	Doubtful design or method (e.g. no hypotheses)
Reliability	[6,8]Intraclass correlation coefficient (ICC)	A relative reliability index incorporating the degree of correlation between TMG measurements (by different raters or in repeated tests). The variance of true TMG scores is set in relation to the variance based on error (i.e. random error and, in some ICC models, systematic bias).	+	ICC ≥ 70%
			−	ICC < 70%
			?	ICC not reported
	[7]Standard error of measurement (SEM)	An index for absolute reliability which quantifies the precision of individual TMG scores in repeated tests, expressed in the units of measurement. The SEM allows calculation of the minimum detectable change (MDC), which is considered the minimum meaningful change in a TMG parameter in repeated testing.	The evaluation cannot be carried out in general but is dependent on the specific application – research question. MDC needs to be evaluated in comparison to the application-specific minimal important change (MIC)	
	[8,9]Coefficient of variation (CV)	An index for absolute reliability which quantifies the variability of individual scores in repeated tests, dimensionless ratio.	+	CV ≤ 10%
			−	CV > 10%
			?	CV not reported
Consistency	[10,11]Cronbach's alpha (α)	An index of the internal consistency of a test or scale.	+	α between ≥ 0.70 and ≤ 0.95
			−	α < 0.70 or > 0.95
			?	Doubtful design or method

1 (Šimundić, 2009); 2 (Guyatt, 2008); 3 (Schisterman et al., 2005); 4 (Shaikh, 2011); 5 (Terwee et al., 2007); 6 (Portney and Watkins, 2009); 7 (Harvill, 1991); 8 (Atkinson and Nevill, 1998); 9 (Stokes, 2009) 10 (Tavakol and Dennick, 2011); 11 (George and Mallery, 2003).

* Adapted from Terwee et al. (2007), Prinsen et al. (2018), and Mokkink et al. (2018); (+) sufficient; (−) insufficient; (?) indeterminate.

methodological quality), moderate (consistent findings in multiple studies of fair methodological quality or in 1 study of good methodological quality), limited (one study of fair methodological quality), conflicting (conflicting findings), unknown (only findings in studies of poor methodological quality). The qualitative ratings of the measurement properties (positive, negative) were based on the results of the quality assessment described in Section 2.8. Disagreements between the reviewers were discussed and clarified by a third author (TS) until a consensus was reached.

3. Results

3.1. Literature search and selection of studies

The systematic search through Medline, Embase and SPORTdiscus databases yielded 373 potential articles, and was supplemented by four papers identified via hand search. The non-specific searches for grey literature produced no relevant hits. After titles and abstracts were reviewed and the duplicates eliminated, 49 full text articles were assessed for eligibility. Nineteen studies met the inclusion criteria for qualitative data analysis Fig. 1 and Table 2. Finally, eleven of those studies were incorporated into the meta-analysis. Fig. 1 shows a flow

Table 2
Demographics and characteristics of studies included.

Study reference, Year, Country	Study Type,	Sample size, sex, age: mean ± SD	Health status, Sport discipline, training intensity	Twitch response, electrode position or IED, ISI	Muscle(s), TMG variables reported	Reference standard and/or comparator	Measurements	Measurement properties
Studies on Diagnostic accuracy tested with QUADAS								
Raeder et al., 2016, Germany	Repeated measures study	23, (14M,24.1 ± 2.0; 9F,25.4 ± 1.9)	Healthy, strength trained athletes (combat sports and intermittent game sports),experienced in strength training ≥2 years, minimum of 2 training units per week	NM, ± 5cm in relation to sensor, NM	Vastus medialis; Dm, Tc, Td, V_{90}	$1RM_{est}$	Three testing sessions: before (baseline) and after an intensified strength training micro-cycle and (Post1) after 3 days of recovery Post4)	Diagnostic Accuracy: PPV, NPV, DE Concurrent validity; criterion: $1RM_{est}$
Wiewelhove et al., 2015, Germany	Interventional study with multiple measures	22, (11M, 22.9 ± 1.9,11F, 23.0 ± 3.4)	Healthy, well-trained team sport athletes, training frequency 5.7 days/week with a mean training volume of 2.5h/day	not adequately described, ± 5cm in relation to sensor, NM	Rectus femoris, biceps femoris (side not mentioned); Dm, Tc	RSA	Three measurements - pre and post of the training program and after 72h of recovery	Diagnostic Accuracy: DE, Youden's index, MR Concurrent validity; criterion: RSA
Wiewelhove et al., 2017, Germany	Interventional study with pre-post design	14, M, 14.9 ± 1.2	Healthy, elite youth tennis player, NM	Individual maximal response, ± 5cm in relation to sensor, 10 s intervals	Rectus femoris (dominant leg); Dm, Tc, V_{10} , V_{90}	CMJ	Two measurements - 24h prior and 24h after a 4-day high-intensity interval training (HIT)	Diagnostic Accuracy: SN, SP, AUC, DE, Youden's index
Studies on Validity tested with CAT								
Harmsen et al., 2018, Germany	Interventional study with multiple measures	10, M, 23.0 ± 3	Healthy, untrained, NM	Individual maximal response, NM, 10s interval	Biceps brachii; Dm, Tc, Td, Ts, Tr	CK, Mb	TMG and biomarkers for EIMD measured prior to EIMD and repeated for the subsequent 3 days	Convergent construct validity: Hypothesis testing
Hunter et al., 2012, UK	Crossover design with multiple measures	19,M, 21.1 ± 4.7	Healthy, NM, NM	Individual maximal response, ± 5cm in relation to sensor, 10s intervals	Biceps brachii; Dm, Tc	MVC, CK	TMG, MVC, biomarkers for EIMD measured prior to EIMD and repeated for the subsequent 6 days	Concurrent validity; criterion: MVC; Convergent construct validity: Hypothesis testing
Sánchez-Sánchez et al., 2018, Spain	Interventional study with multiple measures	20, *M, 25.5 ± 6.1	NM, elite futsal players, NM	25, 50, 75, 100mA, ± 5-6 cm in relation to sensor, NM	Rectus femoris, Biceps femoris; Dm, Tc, Td, Ts, Tr	RSA	TMG, RSA test (7 x 30m)	Concurrent validity; criterion: RSA
Paula Simola et al., 2015/ $2016a_{exp2}$, Germany	Randomized crossover design with multiple measures	14, M, 23.0 ± 1.9	Healthy, experienced in strength training ≥2 years, minimum of 2 training units per week	Individual maximal response, ± 5 cm in relation to sensor, 10s intervals	Rectus femoris both legs (the average value); Dm, V_{10}, V_{90}	MVIC	Measurements: baseline measurements, 24h post intervention following 5 different squat training protocols over 5 weeks	Concurrent validity; criterion: MVIC
Studies on Reliability tested with QAREL								
Ditroilo et al., 2011, Ireland	Descriptive study	16 (14M, 2F) 23.4 ± 4.9 (for reliability evaluation 10 subjects were measured after 48h)	Healthy, NM, students involved in different sports	Individual maximal response, ± 5 cm in relation to sensor, 10s intervals	Biceps femoris; Dm, Tc	nm	TMG, MMT, MIT and EMG: at 3 different knee joint angles 0°/45°/90°	Inter-day reliability: ICC (95% CI), CV (95% CI)
Ditroilo et al., 2013, UK	Crossover design with multiple measures	21,M, 21.3 ± 3.4	Healthy, NM, NM	Individual maximal response, ± 5 cm in relation to sensor, 10s intervals	Gastrocnemius medialis; Dm, Tc, Td, Ts, Tr	nm	Two sessions four weeks apart: TMG baseline, TMG post warm-up, TMG post MVC, TMG post fatigue intervention	Long-term reliability: ICC (95% CI), CV (95% CI), SEM, MDC, MDC (%), Bias
García-García et al., 2018, Spain	Observational study	50, *M, 19.7 ± 2.4	Healthy; elite cyclists, NM	Individual maximal response, NM, 15s intervals	Biceps femoris, Rectus femoris, Vastus lateralis, Vastus medialis; Dm, Tc, Td	nm	Two sessions 20-45 min apart	Intra-day reliability: ICC, CV
Krizja et al., 2008, Slovenia	Observational study	13,M, 30.7 ± 7.4	Healthy, NM, NM	Individual maximal response, ± 5 cm in	Biceps brachii; Dm, Tc, Td, Ts, Tr	nm	30 measurements on each subjects with controlled initial contact pressure	Intra-day reliability: ICC, N-SEM

(*continued on next page*)

Table 2 (*continued*)

Study reference, Year, Country	Study Type,	Sample size, sex, age: mean ± SD	Health status, Sport discipline, training intensity	Twitch response, electrode position or IED, ISI	Muscle(s), TMG variables reported	Reference standard and/or comparator	Measurements	Measurement properties
				relation to sensor, 10s intervals				
Lohr et al., 2018a, Germany	Observational study	24 (13F, 11M) 38 ± 12	Healthy, NM, NM	Individual maximal response, IED ± 3 cm, 10s intervals	Erector spinae; Dm, Tc, Vc	nm	Intra-day: two sessions 3 min apart; Inter-day: post 24h	Intra-day and inter-day reliability: ICC (95% CI), CV (95% CI), SEM, SEM (%), MDC, MDC (%), Bias
Paravlić et al., 2017, Slovenia	Observational study	18 (10M,8F), 30.3 ± 10.3	Healthy, NM, NM	Maximal stimulation at maximal intensity, IED ± 4cm, 10s intervals	Soleus; Dm	nm	Intra-day: two sessions 30 min apart; Inter-day: three sessions within 48h; Inter-rater: two sessions in random order 15 min apart	Intra-day, inter-day and inter-rater reliability: ICC (95% CI), CV, SEM, RE, Bias
Paula Simola et al., 2016a$_{exp1}$, Germany	Descriptive study	20, M, 26.5 ± 6.7.	Healthy, sport students, NM	a) Submaximal response 2 twitches at 40mA b) Individual maximal response, both: ± 5 cm in relation to sensor, 10s intervals	Rectus femoris, biceps femoris, gastrocnemius lateralis; Dm, Tc, Td, Tr, Ts, $V_{10,}$ V_{90}	nm	Inter-day: two testing sessions over one week period	Inter-day reliability: ICC, CV, SEM, Bias
Rey et al., 2012, Spain	Descriptive study	Exp1: 78, M, 26.6 ± 4.4 Exp2 (randomly selected):15,M	Healthy, elite soccer players with a minimum of 4 and a maximum of 15 years of senior soccer activity, NM	Maximal response (50, 75, 100mA), ± 5-6cm in relation to sensor, 10s intervals	Exp1: rectus femoris, biceps femoris Exp2: biceps femoris; Dm, Tc, Td, Ts, Tr	nm	One session: Exp1: comparative evaluation Exp2: estimation of the intra-session reliability	Intra-day reliability: ICC (95% CI)
Rodríguez-Matoso et al., 2010, Spain	Observational study with multiple measures	25, M, 25.7 ± 4.7	NM, students, moderately active	Maximal response (50, 75, 100mA), ± 5cm in relation to sensor, NM	Rectus femoris; Dm, Tc, Td, Ts, Tr	nm	Three measurements: 1.mid-point, 2.and 3. at ± 2cm apart from the first	Intra-day reliability: Cronbach's Alpha
Simunic, 2012, Slovenia	Observational study with multiple measures	10, M, 24.6 ± 3.0	Healthy, NM, NM	Individual maximal response, ± 5cm in relation to sensor, NM	Vastus lateralis, vastus medialis obliquus, biceps femoris; Dm, Tc, Td, Ts, Tr	nm	Three measurements on three occasions over three consecutive days	Inter-day reliability: ICC (95% CI), CV, SEM, RE, Bias
Tous-Fajardo et al, 2010, Spain	Observational study with multiple measures	18, M, 22.9 ± 3.8	Healthy, NM, NM	Individual maximal response, IED ± 3cm and ± 5cm, ≥10s intervals	Vastus medialis; Dm, Tc, Td, Ts, Tr	nm	Four tests separated by rest periods of 3 minutes	Inter-rater and inter-electrode distance reliability: ICC (95% CI), CV, SEM, RE, Bias
Wilson et al., 2018, UK	Within subject with repeated measures	21, (*15M, 6F) 27.0 ± 5.6), for reliability evaluation 10 subjects selected	Healthy, NM, NM	Individual maximal response, ± 2.5 cm in relation to sensor, 1 min intervals	Rectus femoris; Dm	nm	Three consecutive measurements with a rest period of 3 min	Intra-day reliability: ICC, CV

*personal communication; Abbreviations: CK, Creatine Kinase; CMJ, Countermovement jumps; Dm, muscle displacement; EIMD, exercise induced muscle damage; EMG, electromyography; Exp, Experiment; F, female; IED, inter-electrode distance; ISI, inter-stimulus interval; M, male; Mb, Myoglobin; MVC, maximal voluntary contraction; MVIC, maximal voluntary isometric contraction; NM, not mentioned; RSA, repeated sprint ability; RTD, rate of torque development; SD, standard deviation; Tc, time of contraction; Td, delay time; Tr, half-relaxation time; Ts, sustain time; Vc, V_{10}, and V_{90}, velocity of contraction; $1RM_{est}$, estimated one repetition maximum.

chart for the selection process, and in Table 2 the demographics and characteristics of included studies (n = 19) are outlined.

3.2. Study types and measurement properties of included studies

An overview of the evaluated measurement properties in every individual study of the 19 studies included is provided in Table2. The publication by Paula Simola et al. (2016a) comprised two experiments – in the first part the inter-day reliability of TMG was surveyed, and the second experiment investigated the instrument's ability to assess exercise-induced fatigue (i.e. its validity). The data on validity of TMG quoted in the second experiment seem to stem from the same data collection as the data used to analyse the instrument's validity in another included article by the same authors (Paula Simola et al., 2015). Paula Simola et al. (2016a$_{exp2}$) analysed five separate strength training protocols, whereas in the second publication by Paula Simola et al. (2015), the mean values from the five training interventions were averaged for further analysis. Unfortunately, we received no reply from the authors on our enquiry regarding this concern. To avoid a possible over-representation of results stemming from the same dataset, the aforementioned studies (Paula Simola et al., 2015, 2016a$_{exp2}$) will not be rated separately in the further course of the review.

As shown in Table 2, this review comprises 373 healthy volunteers of whom 48 (13%) were women, ranging from 0% (Ditroilo et al., 2013; García-García et al., 2018; Harmsen et al., 2018; Hunter et al., 2012; Krizaj et al., 2008; Paula Simola et al., 2015; Paula Simola et al., 2016a$_{exp1/exp2}$; Rey et al., 2012; Rodríguez-Matoso et al., 2010; Sánchez-Sánchez et al., 2018; Simunic, 2012; Tous-Fajardo et al., 2010; Wiewelhove et al., 2017) to 54% (Lohr et al., 2018a), with average ages ranging from 14.9 ± 1.2 (Wiewelhove et al., 2017) to 38 ± 12 years (Lohr et al., 2018a). The sample size ranged from 10 (Harmsen et al., 2018; Simunic, 2012) to 50 participants (García-García et al., 2018). The included studies focused, with one exception (Lohr et al., 2018a), on muscles of the lower or upper extremities: soleus (Paravlić et al., 2017), gastrocnemius medialis (Ditroilo et al., 2013), gastrocnemius lateralis (Paula Simola et al., 2016a$_{exp1}$), vastus medialis (García-García et al., 2018; Raeder et al., 2016; Simunic, 2012; Tous-Fajardo et al., 2010), vastus lateralis (García-García et al., 2018; Simunic, 2012), rectus femoris (García-García et al., 2018; Paula Simola et al., 2015; Paula Simola et al., 2016a; Rodríguez-Matoso et al., 2010; Sánchez-Sánchez et al., 2018; Wiewelhove et al., 2015; Wiewelhove et al., 2017; Wilson et al., 2018), biceps femoris (Ditroilo et al., 2011; García-García et al., 2018; Paula Simola et al., 2016a$_{exp1}$; Rey et al., 2012; Sánchez-Sánchez et al., 2018; Wiewelhove et al., 2015), biceps brachii (Harmsen et al., 2018; Hunter et al., 2012; Krizaj et al., 2008). The investigation by Lohr et al. (2018a) examined lumbar erector spinae muscles. Three studies reported the inter-electrode distance (IED) (Lohr et al., 2018a; Paravlić et al., 2017; Tous-Fajardo et al., 2010) while the remaining studies specified the position of the electrodes in relation to the sensor, with one exception, Harmsen et al. (2018) have not provided information regarding the precise positioning of the electrodes. The used inter-stimulus interval (ISI) was reported in all studies except Raeder et al. (2016), Rodríguez-Matoso et al. (2010), Sánchez-Sánchez et al. (2018), Simunic (2012), and Wiewelhove et al. (2015). In the following, further information on diagnostic accuracy/validity studies (see Section 3.2.1) and reliability studies (see Section 3.2.2) will be outlined in detail.

3.2.1. Included diagnostic accuracy and validity studies

In case of the six studies that investigated diagnostic accuracy and/or validity the reference standard as well as the fatigue inducing intervention differed between the studies (Table 2). Wiewelhove et al. (2015) and Sánchez-Sánchez et al. (2018) utilized repeated sprint ability (RSA), whereas Raeder et al. (2016) applied estimated 1 repetition maximum (1RM$_{est}$) as reference standard of fatigue, and Wiewelhove et al. (2017) implemented countermovement jump (CMJ) performance. Paula Simola et al. (2015; 2016a$_{exp2}$) and Hunter et al. (2012) used maximal isometric voluntary contraction (MVIC). As indirect measures of exercise induced muscle damage (EIMD) Harmsen et al. (2018) and Hunter et al. (2012) used several biomarkers to assess EIMD. A running-based 6-day high intensity interval training (HIIT), and a 4-day HIIT respectively, was used to induce fatigue in two studies (Wiewelhove et al., 2015; Wiewelhove et al., 2017), whereas Sánchez-Sánchez et al. (2018) implemented Repeated Sprint Ability (RSA) tests to generate fatigue. Raeder et al. (2016) induced muscle fatigue using a 6-day intensified strength training micro-cycle (STM), and five different strength training protocols were conducted in the studies by Paula Simola et al. (2015; 2016a$_{exp2}$). Finally, Hunter et al. (2012) and Harmsen et al. (2018) intended to provoke EIMD by five sets of ten maximal eccentric contractions and six sets of ten eccentric contractions at 110% of the 1RM, respectively.

In case of diagnostic accuracy studies, the measures sensitivity (SN) and specificity (SP) were calculated by Wiewelhove et al. (2017). Positive predictive values (PPV) and negative predictive values (PPV) were assessed by Raeder et al. (2016). One study (Wiewelhove et al., 2017) analysed the area under the curve (AUC), whereas the Youden's index (J) was calculated in two studies (Wiewelhove et al., 2015; Wiewelhove et al., 2017). The misclassification rate (MR) was reported by Wiewelhove et al. (2015). Finally, all aforementioned studies reported the diagnostic effectiveness (DE) (Raeder et al., 2016; Wiewelhove et al., 2015; Wiewelhove et al., 2017). To assess the validity of the studied TMG parameters, criterion related validity (i.e. concurrent validity) was implemented in six studies (Hunter et al., 2012; Paula Simola et al., 2015; Paula Simola et al., 2016a$_{exp2}$; Raeder et al., 2016; Sánchez-Sánchez et al., 2018; Wiewelhove et al., 2015), and construct validity (hypothesis testing, convergent validity) was evaluated by Harmsen et al. (2018) and Hunter et al. (2012).

3.2.2. Included reliability studies

The twelve articles reporting on reliability of TMG assessed different types of reliability (Table 2). Inter-day (Ditroilo et al., 2013; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$; Simunic, 2012) as well as intra-day reliability (Krizaj et al., 2008; Lohr et al., 2018a; Paravlić et al., 2017; Rey et al., 2012; Rodríguez-Matoso et al., 2010; Wilson et al., 2018) was determined. Furthermore, one study focused on inter-day reliability measuring at varying knee joint angles (Ditroilo et al., 2011). Paravlić et al. (2017) and Tous-Fajardo et al., 2010 both assessed inter-rater reliability, additionally the effect of different inter-electrode distance was investigated by Tous-Fajardo et al., 2010 and Wilson et al. (2018). Measurements with differing sensor positioning was investigated by Rodríguez-Matoso et al. (2010).

Eleven studies assessed relative reliability through ICC (Ditroilo et al., 2011; Ditroilo et al., 2013; García-García et al., 2018; Krizaj et al., 2008; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$; Rey et al., 2012; Simunic, 2012; Tous-Fajardo et al., 2010; Wilson et al., 2018), and in one study the authors utilised Cronbach's Alpha as an index of reliability (Rodríguez-Matoso et al., 2010). Absolute reliability was assessed through CV in nine studies (Ditroilo et al., 2011; Ditroilo et al., 2013; García-García et al., 2018; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$; Simunic, 2012; Tous-Fajardo et al., 2010; Wilson et al., 2018), and SEM was calculated in six papers (Ditroilo et al., 2013; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$; Simunic, 2012; Tous-Fajardo et al., 2010). However, in some cases details on methodology of calculation (e.g. ICC model) is lacking of the above mentioned indexes.

3.3. Methodological quality and risk of bias

Each study was assigned to one of the methodological appraisal tools depending on the key aspect and main outcome of the investigation. Three of the studies (Raeder et al., 2016; Wiewelhove et al., 2015;

Table 3
Quality and risk of bias assessment of the included diagnostic accuracy studies, utilized with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool (Whiting et al., 2003).

Author (Year of Publication)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Total	Risk of bias
Raeder et al., 2016	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	N	U	11/14	Low
Wiewelhove et al., 2015	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	U	Y	Y	U	11/14	Low
Wiewelhove et al., 2017	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	11/14	Low

Abbreviations: N, no; U, unclear; Y, yes. *Item 1: was the spectrum of patients representative of those in clinical practice? Item 2: were selection criteria clearly described? Item 3: is the reference standard likely to classify the target condition correctly? Item 4: is the period of time between the reference standard and index test acceptable? Item 5: did the whole sample of patients receive verification using the reference standard? Item 6: did patients receive the same reference standard regardless of the index test result? Item 7: was the reference standard independent of the index test? Item 8: was the execution of the index test described in sufficient detail for replication? Item 9: was the execution of the reference standard described in sufficient detail for replication? Item 10: were the index test results interpreted without knowledge of the reference standard? Item 11: was the reference standard interpreted without knowledge of the results of the index test? Item 12: were the same clinical criteria available when test results were interpreted as would be in clinical practice? Item 13: were uninterpretable/intermediate test results reported? Item 14: were withdrawals from the study explained?

Wiewelhove et al., 2017) focused on diagnostic accuracy and were evaluated via QUADAS. Four papers focusing on validity (Harmsen et al., 2018; Hunter et al., 2012; Paula Simola et al., 2015, 2016a$_{exp2}$; Sánchez-Sánchez et al., 2018) were assessed by CAT and the remaining twelve studies (Ditroilo et al., 2011; Ditroilo et al., 2013; García-García et al., 2018; Krizaj et al., 2008; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$; Rey et al., 2012; Rodríguez-Matoso et al., 2010; Simunic, 2012; Tous-Fajardo et al., 2010; Wilson et al., 2018) set an emphasis on reliability and were evaluated via QAREL.

3.3.1. Quality of diagnostic accuracy studies

The results for the methodological quality and risk of bias scores using QUADAS are gathered in Table 3. All three accuracy studies had a quality score above 10, indicating a high quality and low risk of bias. In one case Item 1 (Was the spectrum of patients representative of those in clinical practice?) was rated 'unclear' because female subjects were not included (Wiewelhove et al., 2017). Due to the unavailability of further details concerning the interpretation of the reference standard and the index test, Item 11 (Was the reference standard interpreted without knowledge of the results of the index test), and furthermore, because of missing information regarding potential withdrawals, Item 14 (Were withdrawals from the study explained) were rated as 'unclear' (Raeder et al., 2016; Wiewelhove et al., 2015; Wiewelhove et al., 2017).

3.3.2. Quality of validity studies

Table 4 shows the methodological quality scores utilizing CAT. Three of the four validity studies being evaluated had a quality score of seven out of nine (Harmsen et al., 2018; Hunter et al., 2012; Paula Simola et al., 2015, 2016a$_{exp2}$), and one study (Sánchez-Sánchez et al., 2018) achieved a score of six out of nine, which all represent a moderate quality as well as a moderate risk of bias. The assessor's qualification or competence (Item 2) was not clarified in any of the studies. In one case (Sánchez-Sánchez et al., 2018), Item 10 (was the execution of the (index) test described in sufficient detail to permit replication of the test?) was rated 'no' because of insufficient information concerning the testing procedure. Item 12 (were withdrawals from the study explained?) was rated 'not applicable', because no information has been provided by the authors.

3.3.3. Quality of reliability studies

The results of the methodological quality assessment of the twelve reliability studies via QAREL are presented in Table 5. Two studies out of twelve were of high quality and low risk of bias with a score of seven out of nine (77.7%; Paravlić et al., 2017), and six out of eight (75.0%; Lohr et al., 2018a), respectively. One investigation achieved a score of six out of nine (66.7%), and was therefore considered to be of moderate quality and moderate risk of bias (Tous-Fajardo et al., 2010). The remaining nine studies had a quality score of 50% or less, and thus, conceivably present a high risk of bias. Only Paravlić et al. (2017) and Lohr et al. (2018a) utilized a representative subject sample (Item 1). Item 2 (Was the test performed by raters who were representative of those to whom the authors intended the results to be applied?) had to be judged as 'unclear' in six out of twelve studies due to insufficient information about the assessor's experience and qualification. In one case (Rodríguez-Matoso et al., 2010), Item 11 (Were appropriate statistical measures of agreement used?) was rated with 'no'. For this reason, the aforementioned study was omitted from the quantitative analysis (see Section 3.4.3).

3.4. Outcome of the studies

3.4.1. Diagnostic accuracy outcome

In total, three studies with low risk of bias evaluated the diagnostic accuracy of two primary (Dm, Tc) and two secondary (V_{10}, V_{90}) TMG

Table 4
Quality and risk of bias assessment of the included validity studies, utilized with the Critical appraisal tool (CAT) (Brink and Louw, 2012).

Author (Year of Publication)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Total	Risk of bias
Harmsen et al, 2018	Y	N	Y	N/A	N/A	N/A	Y	N/A	Y	Y	Y	U	Y	7/9	Moderate
Hunter et al., 2012	Y	N	Y	N/A	N/A	N/A	Y	N/A	Y	Y	Y	U	Y	7/9	Moderate
Paula Simola et al., 2015/2016a$_{exp2}$	Y	N	Y	N/A	N/A	N/A	Y	N/A	Y	Y	Y	U	Y	7/9	Moderate
Sánchez-Sánchez et al., 2018	Y	N	Y	N/A	N/A	N/A	Y	N/A	Y	N	Y	U	Y	6/9	Moderate

Abbreviations: N, no; U, unclear; Y, yes; N/A, not applicable. *Item 1: If human subjects were used, did the authors give a detailed description of the sample of subjects used to perform the (index) test? Item 2: Did the authors clarify the qualification, or competence of the rater(s) who performed the (index) test? Item 3: Was the reference standard explained? Item 4: If interrater reliability was tested, were raters blinded to the findings of other raters? Item 5: If intrarater reliability was tested, were raters blinded to their own prior findings of the test under evaluation? Item 6: Was the order of examination varied? Item 7: If human subjects were used, was the time period between the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests? Item 8: Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time interval between repeated measures? Item 9: Was the reference standard independent of the index test? Item 10: Was the execution of the (index) test described in sufficient detail to permit replication of the test? Item 11: Was the execution of the reference standard described in sufficient detail to permit its replication? Item 12: Were withdrawals from the study explained? Item 13: Were the statistical methods appropriate for the purpose of the study?

Table 5
Quality and risk of bias assessment of the included diagnostic reliability studies, utilized with the Quality Appraisal of Reliability Studies (QAREL) checklist (Lucas et al., 2010).

Author (Year of Publication)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Total	Risk of bias
Ditroilo et al., 2011	U	U	N/A	U	N/A	N/A	U	Y	Y	Y	Y	4/8	High
Ditroilo et al., 2013	U	U	N/A	U	N/A	N/A	U	U	Y	Y	Y	3/8	High
García-García et al., 2018	U	Y	N/A	U	N/A	N/A	U	U	Y	Y	Y	4/8	High
Krizaj et al., 2008	U	U	N/A	U	N/A	N/A	U	U	Y	Y	Y	3/8	High
Lohr et al., 2018a	Y	Y	N/A	U	N/A	N/A	U	Y	Y	Y	Y	6/8	Low
Paravlić et al., 2017	Y	Y	Y	U	N/A	N/A	U	Y	Y	Y	Y	7/9	Low
Paula Simola et al., 2016a$_{exp1}$	U	U	N/A	U	N/A	N/A	U	U	Y	Y	Y	3/8	High
Rey et al., 2012	U	Y	N/A	U	N/A	N/A	U	U	Y	Y	Y	4/8	High
Rodríguez-Matoso et al., 2010	U	U	N/A	U	N/A	N/A	U	U	Y	Y	N	2/8	High
Simunic, 2012	U	Y	N/A	U	N/A	N/A	U	U	Y	Y	Y	4/8	High
Tous-Fajardo et al., 2010	U	Y	Y	U	N/A	N/A	U	Y	Y	Y	Y	6/9	Moderate
Wilson et al., 2018	U	U	N/A	U	N/A	N/A	U	U	Y	Y	Y	3/8	High

Abbreviations: N, no; U, unclear; Y, yes; N/A, not applicable. *Item 1: Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? Item 2: Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? Item 3: Were raters blinded to the findings of other raters during the study? Item 4: Were raters blinded to their own prior findings of the test under evaluation? Item 5: Were raters blinded to the results of the reference standard for the target disorder (or variable) being evaluated? Item 6: Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? Item 7: Were raters blinded to additional cues that were not part of the test? Item 8: Was the order of the examination varied? Item 9: Was the time interval between repeated measurements compatible with the stability (or theoretical stability) of the variable being measured? Item 10: Was the test applied correctly and interpreted appropriately? Item 11: Were appropriate statistical measures of agreement used?

measures (Raeder et al., 2016; Wiewelhove et al., 2015; Wiewelhove et al., 2017). The results are outlined in detail in Table 6. Table 7 shows the quality of the assessed measurement properties (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2007). The TMG parameter Dm was rated insufficient on diagnostic accuracy (SN, SP, PPV, NPV, DE, MR, *J*, AUC) in all three studies. The same applies for the time-based parameter Tc, being rated insufficient by two of the aforementioned investigations (Wiewelhove et al., 2015; Wiewelhove et al., 2017). The assessed accuracy measures were: SN, SP, DE, MR, *J*, and AUC. Only Wiewelhove et al. (2017) evaluated the secondary TMG variable V_{10}. All surveyed data on diagnostic accuracy (SN, SP, DE, MR, J, and AUC) were rated insufficient. Finally, the accuracy of V_{90} was assessed by Raeder et al. (2016) and Wiewelhove et al. (2017). This parameter was rated insufficient as well, with one exception: Raeder et al. (2016) assessed PPV, inter alia, and compared the measurements at several time points. In one case (changes between pre- and post-training after a 6-day HIIT), PPV of V_{90} reached 80% and was therefore rated sufficient. However, taking all available data together, the tested TMG parameters demonstrated questionable overall diagnostic accuracy to discriminate between fatigued and recovered athletes.

3.4.2. Validity outcome

The results of the six studies on validity are summarized in Tables 8 and 9. In terms of concurrent criterion related validity, the assessed TMG parameters were rated insufficient due to the fact that none of the analysed measures (Dm, Tc, Td, V_{10}, and V_{90}) exceeded a correlation of ≥ 0.70 with the criterion measure of fatigue (Hunter et al., 2012; Paula Simola et al., 2015; Raeder et al., 2016; Sánchez-Sánchez et al., 2018; Wiewelhove et al., 2015), with one exception for the parameter Dm in the study by Paula Simola et al. (2016a$_{exp2}$). Here, changes between pre- and post-training in Dm correlated ($r = 0.705$) with the criterion measure (MVIC) after one out of five training protocols and was therefore rated sufficient for this protocol. Two studies with moderate risk of bias (Harmsen et al., 2018; Hunter et al., 2012) evaluated convergent construct validity of the variable Dm. In case of Harmsen et al. (2018), convergent construct validity was sufficient, whereas the validity measures determined by Hunter et al. (2012) were rated insufficient. The convergent construct validity of the parameters Tc and Td – assessed only by Harmsen et al. (2018) – was rated insufficient in both cases.

3.4.3. Reliability outcome

The results on relative and absolute reliability of TMG measures (Dm, Tc, Td, Vc, V_{10}, and V_{90}) are gathered in Table 10. With regard to relative reliability, one of the original twelve reliability studies that were included in the qualitative assessment had to be excluded from the quantitative assessment via meta-analysis because of the use of Cronbach's alpha instead of ICC (Rodríguez-Matoso et al., 2010).

Eleven studies (one inter-rater and eleven test-retest reliability) evaluated the relative reliability (i.e. ICC) of the spatial parameter Dm. The time-dependent measures Tc and Td were analysed in nine (one inter-rater and nine test-retest), and in seven studies (one inter-rater and seven test-retest reliability), respectively. The relative reliability results for the parameters Dm, Tc, and Td were summarized quantitatively in a meta-analysis. The evaluated heterogeneity (τ^2) for the TMG measures was as follows: Dm: 0.01%, Tc: 0.04%, and Td: 0.07%, respectively. Thus, the heterogeneity was so small that the reliabilities of each study could be pooled. The point estimates with respective CIs for the reliabilities of Dm, Tc, and Td were 0.98 (0.96; 0.99), 0.95 (0.93; 0.98), and 0.91 (0.88; 0.95), respectively. The forest-plots illustrate these results (Figs. 2–4). These values indicate an excellent relative reliability for all three assessed TMG parameter (Dm, Tc, and Td). As mentioned above, the study by Rodríguez-Matoso et al. (2010) had to be excluded from the meta-analysis due to the use of inappropriate statistical measures (Cronbach's alpha). Nevertheless, the results of this study were also rated sufficient for the analysed parameters (Dm, Td, Tc), and thus support the aforementioned outcomes. As described in Section 3.3.3), the secondary parameter velocity of contraction (Vc, V_{10}, V_{90}) had been assessed by only two test-retest studies, one study with low risk of bias (Lohr et al., 2018a), and one with high risk of bias (Paula Simola et al., 2016a$_{exp1}$). Because of the scant data available and varying methods of calculation between the two studies, a meta-analysis was not performed for this TMG measure. However, in both investigations, the velocity parameters showed excellent relative reliability – Vc: ICC scores ≥ 0.97 (Lohr et al., 2018a), V_{10}/V_{90}: ICC scores ≥ 0.92 Paula Simola et al. (2016a$_{exp1}$), and were rated sufficient.

Absolute reliability, i.e. CV, was assessed in nine studies for the parameter Dm (two inter-rater and nine test-retest). The measures Tc and Td were analysed in seven (one inter-rater and seven test-retest), and in five studies (one inter-rater and five test-retest), respectively. The results for the quality of CV are outlined in Table 11. In four of the nine studies, Dm was rated sufficient (García-García et al., 2018; Simunic, 2012; Tous-Fajardo et al., 2010; Wilson et al., 2018), while in

Table 6
Diagnostic accuracy of tensiomyographic measures to discriminate between fatigued and non-fatigued subjects.

Author, Year	Experimental test/ criterion measure/ subjects	Muscle(s), TMG variables	SN (%)	SP (%)	PPV (%)	NPV (%)	AUC (95% CI)	DE (%)	Youden's index	MR (%)	Conclusion by the authors
Raeder et al., 2016	Accuracy of TMG parameters to differentiate between fatigued and recovered athletes in relation to the criterion measure: $1RM_{est}$, n = 23	Vastus medialis (side not mentioned), Dm, V_{90}	–	–	Dm: pre-post1 = 75.0, pre-post4 = 50.0; V_{90}: pre-post1 = 80.0, pre-post4 = 66.7	Dm: pre-post1 = 42.9. pre-post4 = 66.7; V_{90}: pre-post1 = 50.0, pre-post4 = 69.2	–	Dm: pre-post1 = 65.2, pre-post4 = 62.5; V_{90}: pre-post1 = 69.6, pre-post4 = 68.8	–	–	TMG markers showed insufficient accuracy to discriminate between fatigued and recovered athletes in relation to the criterion measure
Wiewelhove et al., 2017	Accuracy of TMG parameters to detect muscle fatigue in relation to the criterion measure: CMJ, n = 14	Rectus femoris (dominant side), Dm, Tc, V_{10}, V_{90}	Dm: 33.3, Tc: 33.3, V_{10}: 33.3, V_{90}: 33.3	Dm: 50.0, Tc: 50.0, V_{10}: 50.0, V_{90}: 50.0	–	–	Dm: 0.46 (0.15–0.77), Tc: 0.29 (0.03–0.55), V_{10}: 0.71 (0.27–1.00), V_{90}: 0.37 (0.10–0.65)	Dm: 35.7, Tc: 35.7, V_{10}: 35.7, V_{90}: 35.7	Dm: 0.17 Tc: 0.17, V_{10}: 0.17, V_{90}: 0.17	–	TMG markers were not sensitive enough to detect muscular performance changes in elite youth athletes
Wiewelhove et al., 2015	Accuracy of TMG parameters to detect muscle fatigue and recovery in relation to the criterion measure: RSA, n = 22	Rectus femoris, biceps femoris (side not mentioned), Dm, Tc	–	–	–	–	–	**RF:** Dm: pre-post1 = 50.0, pre-post2 = 66.7; Tc: pre-post1 = 68.2, pre-post2 = 40.0 **BF:** Dm: pre-post1 = 50.0, pre-post2 = 60.0; Tc: pre-post1 = 54.5, pre-post2 = 40.0	**RF:** Dm: pre-post1 = 0.04, pre-post2 = 0.30; Tc: pre-post1 = 0.38, pre-post2 = 0.21 **BF:** Dm: pre-post1 = 0.11, pre-post2 = 0.20; Tc: pre-post1 = 0.03, pre-post2 = 0.23	**RF:** Dm: pre-post1 = 50.0, pre-post2 = 33.3; Tc: pre-post1 = 31.8, pre-post2 = 60.0 **BF:** Dm: pre-post1 = 50.0, pre-post2 = 40.0; Tc: pre-post1 = 45.5, pre-post2 = 60.0	TMG markers showed insufficient accuracy to discriminate between fatigued and recovered athletes in relation to the criterion measure

Abbreviations: AUC, Area under the curve; BF, Biceps femoris; CI, confidence interval; DE, diagnostic effectiveness; MR, misclassification rate; NPV, negative predictive value, PPV, positive predictive value; RF, Rectus femoris; RSA, repeated sprint ability; SN, Sensitivity; SP, Specificity; Tc, contraction time; V_{10} and V_{90}, velocity of contraction; $1RM_{est}$, estimated one repetition maximum.

Table 7
Quality of measurement properties for diagnostic accuracy.

TMG Variables	Study reference	Diagnostic accuracy: Measurement property	Diagnostic accuracy: rating	Risk of bias
Dm	Raeder et al., 2016	PPV	−/−	Low
		NPV	−/−	
		DE	−/−	
	Wiewelhove et al., 2015	DE	−/−/−/−	Low
		MR	−/−/−/−	
		J	−/−/−/−	
	Wiewelhove et al., 2017	SN	−	Low
		SP	−	
		DE	−	
		AUC	−	
		J	−	
Tc	Wiewelhove et al., 2015	DE	−/−/−/−	Low
		MR	−/−/−/−	
		J	−/−/−/−	
	Wiewelhove et al., 2017	SN	−	Low
		SP	−	
		DE	−	
		AUC	−	
		J	−	
V_{10}	Wiewelhove et al., 2017	SN	−	Low
		SP	−	
		DE	−	
		AUC	−	
		J	−	
V_{90}	Raeder et al., 2016	PPV	+/−	Low
		NPV	−/−	
		DE	−/−	
	Wiewelhove et al., 2017	SN	−	Low
		SP	−	
		DE	−	
		AUC	−	
		J	−	

(+) sufficient; (−) insufficient; (?) indeterminate; Abbreviations: AUC, area under the curve; DE, diagnostic effectiveness; Dm, muscle displacement; *J*, Youden's index; MR, misclassification rate; NPV, negative predictive value; PPV, positive predictive value; *r*, correlation coefficient; SN, sensitivity; SP, specificity; Tc, time of contraction; V_{10}, V_{90}, velocity of contraction.

the study by Ditroilo et al. (2011), CV was > 10%, and therefore rated insufficient. In the remaining four publications (Ditroilo et al., 2013; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$) the results for Dm varied between protocols within each study. In the studies by Ditroilo et al. (2013) and Lohr et al. (2018a), Dm was rated insufficient in three out of four protocols. In case of the investigation by Paravlić et al. (2017), Dm was found to be sufficient in two out of three protocols. Finally, the ratings for Dm in the study by Paula Simola et al. (2016a$_{exp1}$) varied depending on the examined muscle. Dm was rated sufficient for the examination of rectus femoris, whereas the parameter was found to be insufficient in case of biceps femoris and gastrocnemius lateralis. The time-dependent variables Tc and Td were rated sufficient (Ditroilo et al., 2013; García-García et al., 2018; Lohr et al., 2018a; Paula Simola et al., 2016a$_{exp1}$; Simunic, 2012; Tous-Fajardo et al., 2010), with one exception for the parameter Tc as assessed by Ditroilo et al. (2011). Here, the parameter was rated insufficient. The absolute reliability of the TMG-derived variables Vc, V_{10}, and V_{90} was investigated, in each case, by one test-retest study. The parameter Vc (CV < 8%) was rated sufficient (Lohr et al., 2018a), whereas the ratings for V_{10} and V_{90} (Paula Simola et al., 2016a$_{exp1}$) again differed depending on the examined muscle. The variables were rated sufficient in case of rectus femoris and biceps femoris (CV < 10%), while the rating was insufficient for the evaluation of gastrocnemius lateralis (CV > 10%).

3.5. Strength of evidence

A summary of the evidence synthesis for each individual TMG variable is provided in Table 12. With respect to the TMG parameters' overall diagnostic accuracy, the evidence ranged from moderate negative (V_{10}) to strong negative (Dm, Tc, and V_{90}). The determined evidence for the parameters' accuracy are based on small total numbers of participants across the TMG parameters, ranging from 14 to 59 volunteers, with a notably low share of female participants (range: 0–34%).

In terms of concurrent criterion validity, the evidence was found to be limited negative for the variables Td, and V_{10}, and moderate negative for the variables Tc and V_{90}. Finally, Dm was rated strong negative. The evidence for convergent construct validity (Dm) was rated conflicting, whereas the parameters Td and Tc were rated limited negative. Again, it should be mentioned, that the data rely upon relatively small total numbers of subjects (range: 10–98), with a remarkably low share of female participants (range: 0–26%).

Regarding the TMG parameters' relative reliability, the findings are based on the pooled data from the meta-analysis for the parameters Dm, Tc, and Td, and the synthesized evidence from the two studies examining the secondary TMG parameters. The evidence from the meta-analysis was strong positive for all included parameters, with total sample sizes ranging from 147 (Td) to 209 (Dm). Finally, the evidence for the secondary TMG parameters (total sample sizes: 20–24) ranged from unknown (V_{10}, V_{90}; due to the high risk of bias of the only study assessing these parameters) to moderate positive (Vc). Again, the percentage of female participants was predominantly low (0–13%), with one exception for the parameter Vc (54%).

In contrast to the findings of relative reliability, the evidence regarding TMG's absolute reliability varied. Depending on the parameter, the evidence was found to be unknown (V_{10}, V_{90}; see above), conflicting (Dm), limited positive (Td), or moderate positive (Tc, Vc). The assessed results are based on total numbers of subjects ranging from 20 (V_{10}, V_{90}) to 181 (Dm), again with a predominantly low share of women (range: 0–54%).

4. Discussion

4.1. Main findings

We conducted a comprehensive systematic review on the diagnostic accuracy, validity, and reliability of the three primary TMG measures muscle displacement (Dm), contraction time (Tc), and delay time (Td), and the secondary TMG-derived parameter velocity of contraction (Vc, V_{10}, V_{90}). In case of relative reliability, also a quantitative assessment (meta-analysis) was carried out for the three primary TMG parameters. The methodological quality was assessed for each of the 19 studies that were included. The three studies focusing on diagnostic accuracy were evaluated via QUADAS (Whiting et al., 2003), the four validity studies were assessed using CAT (Brink and Louw, 2012), and the remaining twelve studies, which focused on reliability, were rated by means of QAREL (Lucas et al., 2010).

The main findings of the present review suggest that robust evidence for diagnostic accuracy of TMG measures (Dm, Tc, V_{10}, and V_{90}) to discriminate between fatigued and recovered healthy subjects has yet to be determined (Raeder et al., 2016; Wiewelhove et al., 2015; Wiewelhove et al., 2017). A similar picture emerges with regard to criterion referenced validity (Hunter et al., 2012; Paula Simola et al., 2015; Paula Simola et al., 2016a$_{exp2}$; Raeder et al., 2016; Wiewelhove et al., 2015). Construct validity was found to be conflicting for the variable Dm (Harmsen et al., 2018; Hunter et al., 2012), or limited negative for Tc and Td (Harmsen et al., 2018).

Moreover, as demonstrated in the meta-analysis of the reliability studies, the relative reliability for the analysed TMG parameters was high to excellent (Td 0.91, Tc 0.95, Dm 0.98) (Ditroilo et al., 2011;

Table 8
Validity of tensiomyographic measures to assess muscle fatigue.

Author, Year	Muscle(s), TMG variables, subjects	Data analysis		Conclusions by the authors
		Outcome tested	Data analysis / results	
Harmsen et al., 2018	Biceps brachii, Dm, n = 10	Validity of TMG Dm in detecting changes in MVC following EIMD	Correlation's coefficient for % changes from pre-Ex for TMG and EIMED biomarkers: Post-Ex: CK ($r = -0.74^{*}$), Mb ($r = -0.43$); 20 min CK ($r = -0.66^{*}$), Mb ($r = -0.62$); 2 h CK ($r = -0.42$), Mb ($r = -0.19$); 24 h CK ($r = -0.62$), Mb ($r = -0.73^{*}$); 48 h CK ($r = -0.95^{**}$), Mb ($r = -0.87^{**}$); 72 h CK ($r = -0.79^{**}$), Mb ($r = -0.66^{**}$). Tc vs CK: no large correlation; Td vs CK: no correlation	TMG data correlate with muscle damage markers after eccentric exercise of the elbow flexors of untrained males
Hunter et al., 2012	Biceps brachii, Dm, n = 19	Validity of TMG Dm in detecting changes in MVC following EIMD	MVC vs TMG Pearson's correlation for % change of MVC vs TMG Dm from day 0 (pre damage) until day 1,2,3,4: $r = 0.55^{*} \pm 0.20$ (95% CI: 0.12–0.96), $r = 0.56^{*} \pm 0.21$ (0.11–1.02), $r = 0.68^{*} \pm 0.19$ (0.27–1.09), $r = 0.67^{*} \pm 0.27$ (0.092–1.24) For days 5 and 6 the relationship decreased by showing $r = 0.13 \pm 0.34$ (0.57–0.84) and $r = 0.36^{*} \pm 0.16$ (0.011–0.69), respectively. EIMD biomarker vs TMG Pearson's correlation for % change of CK vs TMG Dm from day 0 (pre damage) until day 1, 3, and 6: $r = 0.15$, $r = -0.47^{*}$, $r = -0.59^{**}$	TMG Dm was effective in detecting muscle damage
Paula Simola et al., 2015/2016a$_{exp2}$	Rectus femoris, Dm, V_{10}, V_{90}, n = 14	2015: Pearson's correlation coefficients for changes of MVIC and TMG parameters between baseline and post-train after all squat protocols (averaged) 2016$_{exp2}$: Correlation for changes of MVIC and TMG parameters between baseline and post-train after 5 squat protocols	Correlation for changes of MVIC and TMG parameter between baseline and post-train measurements Dm ($r = 0.64$), V_{10} ($r = 0.67$), and V_{90} ($r = 0.66$) Correlation between changes of pre- and post-train in Dm, V_{10}, V_{90} with changes in MVIC after MS: $r = 0.169, 0.134, 0.59$; after DS: $r = 0.346, 0.454, 0.482$; after FW: $-0.234, -0.276, -0.270$; after EO: $r = 0.705^{*}, 0.699^{*}, 0.695^{*}$; after PL: $r = 0.389, 0.406, 0.433$	Some TMG parameters are sensitive to changes in muscle force Changes between pre- and post-train in Dm, V_{10} and V_{90} correlated with changes in MVIC only after the EO protocol
Raeder et al., 2016	Vastus medialis, Dm, V_{90}, n = 23	Correlational analyses to evaluate relationships between changes in TMG markers and changes in $1RM_{est}$ after 6-day intensified strength training micro-cycle	Correlational analyses revealed no significant correlations ($p > .05$) between changes in $1RM_{est}$ and TMG markers of fatigue and recovery	There is no significant correlation between changes in $1RM_{est}$ and TMG markers
Sánchez-Sánchez et al., 2018	Rectus femoris, Biceps femoris, Dm, Td, Tc, n = 20	Pearson's correlation coefficient was calculated between the results of the RSA test and the TMG variables	Correlation between the RSA change and the percentage change in the TMG parameters from pre to post of the RSA test. RF: Dm ($r = -0.485^{*}$), Td ($r = -0.043$), Tc ($r = -0.263$); BF: Dm ($r = 0.018$), Td ($r = 0.578^{**}$), Tc ($r = 0.469^{*}$)	The findings suggest that muscle mechanical variables have a significant relationship with the performance during a RSA test
Wiewelhove et al., 2015	Rectus femoris, biceps femoris, Dm, Tc, n = 22	Multiple regression analysis was used to assess relationships between changes in TMG markers and RSA after 6-day running-based high-intensity interval micro-cycle	Multiple regression analysis revealed no significant correlations ($p > 0.05$) between changes in RSA and TMG markers of fatigue and recovery	There is no significant correlation between changes in RSA and TMG markers

Abbreviations: Dm, muscle displacement; DS, Drop sets; EIMD, exercise-induced muscle damage; EO, Eccentric overload; FW, Flywheel; MVC, maximal voluntary contraction; MVIC, maximal voluntary isometric contraction; MS, multiple sets; PL, Plyometrics; RSA, repeated sprint ability; Tc, time of contraction; V_{10} and V_{90}, velocity of contraction; $1RM_{est}$, one estimated repetition maximum.

* $p = < 0.05$.

** $p = < 0.01$.

Table 9
Quality of measurement properties for validity.

		Validity				Risk of bias
		Criterion		Construct		
TMG Variables	Study reference	Measurement property	rating	Measurement property	rating	
Dm	Harmsen et al., 2018			*r*	+	Moderate
	Hunter et al., 2012	*r*	−/−/−/−/−/−	*r*	−	Moderate
	Paula Simola et al., 2015/2016a$_{exp2}$	*r*	−/−/−/−/+/−			Moderate
	Raeder et al., 2016	*r*	−			Low
	Sánchez-Sánchez et al., 2018	*r*	−/−			Moderate
	Wiewelhove et al., 2015	*r*	−			Low
Td	Harmsen et al., 2018			*r*	−	Moderate
	Sánchez-Sánchez et al., 2018	*r*	−/−			Moderate
Tc	Harmsen et al., 2018			*r*	−	Moderate
	Sánchez-Sánchez et al., 2018	*r*	−/−			Moderate
	Wiewelhove et al., 2015	*r*	−			Low
V_{10}	Paula Simola et al., 2015/2016a$_{exp2}$	*r*	−/−/−/−/−/−			Moderate
V_{90}	Paula Simola et al., 2015/2016a$_{exp2}$	*r*	−/−/−/−/−/−			Moderate
	Raeder et al., 2016	*r*	−			Low

(+) sufficient; (−) insufficient; (?) indeterminate; *Abbreviations:* AUC, area under the curve; DE, diagnostic effectiveness; Dm, muscle displacement; *J*, Youden's index; MR, misclassification rate; NPV, negative predictive value; PPV, positive predictive value; *r*, correlation coefficient; SN, sensitivity; SP, specificity; Tc, time of contraction; V_{10}, V_{90}, velocity of contraction.

Ditroilo et al., 2013; García-García et al., 2018; Krizaj et al., 2008; Lohr et al., 2018a; Paravlić et al., 2017; Paula Simola et al., 2016a$_{exp1}$; Rey et al., 2012; Simunic, 2012; Tous-Fajardo et al., 2010; Wilson et al., 2018). Our results demonstrate strong positive evidence to support relative reliability for the analysed TMG variables. In terms of absolute reliability the results were less unequivocal for the primary TMG measures, ranging from conflicting (Dm) to moderate positive (Tc). In case of the indirect measures of contraction speed the evidence ranged from unknown (V_{10}, V_{90}) to moderate positive (Vc) for both relative and absolute reliability.

Finally, the data collected here suggest that there is a crucial need to implement higher methodological standards for the conduct and reporting of TMG studies. This particularly applies for the included reliability studies – only two out of twelve studies were of high quality, with a low potential risk of bias considering the pre-established definition of quality ranking (Lohr et al., 2018a; Paravlić et al., 2017). In the following, the findings will be discussed in more detail.

4.1.1. Diagnostic accuracy and validity

One objective of this study was to review available data on the accuracy and validity of TMG parameters to differentiate between fatigued and non-fatigued individuals. In case of diagnostic accuracy studies, TMG failed to detect post-interventional muscular performance alterations. The evaluated overall evidence for criterion related validity ranged from limited negative to moderate negative, since the prevailing majority of the analysed parameters (Dm, Tc, V_{10}, and V_{90}) did not exceed a correlation of ≥ 0.70 with the reference criterion. In some cases, however, there were significant correlations between changes in the considered gold standard of fatigue (i.e. maximal voluntary contraction) and changes in the spatial TMG measure Dm following EIMD (Hunter et al., 2012), or in the TMG variables Dm, V_{10}, and V_{90} subsequent to an eccentric overload strength intervention protocol (Paula Simola et al., 2016a$_{exp2}$). The conflicting evidence assessed for construct validity (Dm) failed to underpin this slight positive trend sufficiently.

Apart from general methodological shortcomings (see Sections 4.2. and Section 4.3 below), these mostly unsatisfactory results might be specifically attributable to issues related to the experimental setup. The nature of exercise-induced muscle fatigue is very complex, and its characteristics are influenced (amongst others) by the task being performed – such as the type of contraction as well as its duration and intensity (Millet and Lepers, 2004). For example, it could be shown that isometric, concentric, and eccentric muscle activity differ in their neuromuscular fatigue profile (Kay et al., 2000). Another point to note is that the participant's training background (e.g. endurance-trained vs. power-trained subjects) affects functional structures of the neuromuscular system (Garrandes et al., 2007). Looking at the experimental setup of the studies by Wiewelhove et al. (2015; 2017), the authors implemented running-based interventions (4-day HIIT, 6-day HIIT), and as criterion measures of fatigue, they used countermovement jump performance and repeated sprint ability. Since these skills or tasks involve multiple well-defined and coordinated muscle groups (Zarrouk et al., 2012), it is questionable whether the interventions and criterion measures of fatigue chosen for these studies were potent enough to have an adequate impact on TMG measures. In this context, one should also keep in mind that TMG is a method for the evaluation of isolated muscular structures (Simunic, 2012). This assumption is backed up by the validity investigation of Paula Simola et al. (2015; 2016a$_{exp2}$) who assessed neuromuscular function after five different strength training protocols. The authors only found positive correlations between the reduction in muscle force capacity and alterations in the TMG parameters Dm, V_{10}, and V_{90} subsequent to the eccentric overload protocol (Paula Simola et al., 2016a$_{exp2}$; cf. Tables 8 and 9). This observation supports the assumption that the effect on TMG measures is dependent on the training load and on its duration (Paula Simola et al., 2016b). Consequently, routines that include eccentric exercise might have a greater impact on TMG indices, especially when applied to strength-trained athletes. Consistent with these findings, Hunter et al. (2012) – who assessed local eccentric EIMD of the elbow flexors via TMG – found that post-interventional changes of the TMG parameter Dm significantly correlated with reductions in muscle force capacity (Torque vs TMG Dm ranged from $r = 0.55$ [± 0.20] to $r = 0.67$ [± 0.27]). Immediate alterations of TMG Dm after performing eccentric EIMD of the elbow flexors were also observed in a recently published study (Franz et al., 2017), and thus support the notion that local eccentric exercise may lead to higher accuracy of TMG measures.

Here it should be mentioned that muscle fatigue, which usually requires a short regeneration period, ought to be distinguished from muscle damage, which was investigated by Hunter et al. (2012) and Harmsen et al. (2018). Muscle damage is characterized by a prolonged phase of regeneration (Finsterer, 2012). Moreover, muscle damage corresponds with structural changes (e.g. damage to sarcomeres in myofibrils, damage to the excitation-contraction coupling system), whereas fatigue is considered as a functional sign (Allen et al., 2005;

Table 10
Reliability of tensiomyographic measures.

					Data analysis						
Author, Year	Type of Reliability, subjects	Muscle(s)	TMG variables		Relative Reliability	Absolute Reliability					
Ditroilo et al., 2011	Inter-day reliability (48 h, varying knee joint angles), n = 10			Angle (°)	ICC, Model Yes, (95% CI)	CV, (95% CI)					
		BF	Dm	0°	0.82 (0.42–0.95)	19.8 (11.9–27.6)					
				45°	0.57 (−0.05–0.87)	19.7 (11.2–28.1)					
				90°	−0.57 (−0.87–0.05)	43.1 (24.4–61.7)					
			Tc	0°	0.82 (0.42–0.95)	16.5 (8.9–24.0)					
				45°	0.62 (0.03–0.89)	20.5 (10.7–30.3)					
				90°	−0.40 (17.0–49.7)	33.3 (−0.81–0.27)					
Ditroilo et al., 2013	Inter-day reliability (4 weeks), n = 21			Time point of measurement	ICC, Model Yes, (95% CI)	CV, (95% CI)	SEM	MDC	MDC %	Bias	
		GM	Dm	M1	0.86 (0.68–0.94)	14.8 (10.4–19.3)	0.30	0.84	22.07	$p < .05$	
				M2	0.91 (0.79–0.96)	11.1 (7.0–15.1)	0.26	0.72	18.19	$p < .01$	
				M3	0.92 (0.82–0.97)	10.1 (5.0–15.2)	0.21	0.58	17.13	ns	
				M4	0.95 (0.87–0.98)	8.0 (5.0–11.0)	0.19	0.51	15.05	ns	
			Tc	M1	0.62 (0.27–0.82)	9.4 (6.3–12.5)	1.36	3.77	15.79	ns	
				M2	0.63 (0.28–0.82)	9.1 (6.4–11.8)	1.36	3.77	14.95	ns	
				M3	0.62 (0.26–0.82)	8.1 (5.4–10.8)	1.18	3.28	13.70	ns	
				M4	0.92 (0.82–0.97)	3.8 (2.3–5.3)	0.61	1.68	7.33	ns	
			Td	M1	0.60 (0.24–0.82)	9.2 (5.2–12.9)	1.34	3.72	16.41	ns	
				M2	0.56 (0.18–0.80)	8.2 (4.2–12.3)	1.52	4.21	19.05	ns	
				M3	0.62 (0.27–0.83)	7.8 (3.7–11.9)	1.27	3.53	17.15	ns	
				M4	0.62 (0.26–0.82)	7.0 (3.2–10.8)	1.16	3.20	14.98	ns	
García-García et al., 2018	Intra-day reliability; 2 sessions 20–45 min apart, n = 50	BF, RF, VL, VM randomly selected			ICC (0.95% CI), Model Yes	CV					
			Dm		0.99 (0.98–1.00)	1.6					
			Tc		0.97 (0.94–0.99)	4.0					
			Td		0.91 (0.80–0.96)	3.4					
Krizaj et al., 2008	Short-term test–retest reliability, n = 13				ICC, Model Yes	N-SEM					
		BB	Dm		0.98	1.23					
			Tc		0.97	0.48					
			Td		0.94	0.48					
Lohr et al., 2018a	Intra-day and inter-day reliability, 2 sessions, n = 24			Time points of measurement	ICC, Model Yes, (95% CI)	CV (95% CI)	SEM	SEM %	MDC	MDC%	Bias
		ES	Dm right	M1	0.95 (0.88–0.98)	10.2 (5.4–14.9)	0.54	13.2	1.50	36.7	ns
			Dm left	M1	0.94 (0.87–0.97)	12.2 (7.0–17.4)	0.52	13.3	1.50	38.5	ns
			Dm right	M2	0.96 (0.90–0.98)	12.3 (8.4–16.2)	0.51	12.3	1.41	34.1	ns
			Dm left	M2	0.96 (0.92–0.98)	9.7 (6.6–12.9)	0.41	10.6	1.14	29.5	ns
			Tc right	M1	0.88 (0.74–0.95)	4.4 (2.6–6.3)	1.40	7.3	3.90	20.4	ns
			Tc left	M1	0.81 (0.60–0.91)	5.2 (2.3–8.2)	1.72	8.9	4.80	24.7	ns
			Tc right	M2	0.80 (0.59–0.91)	6.9 (4.3–9.6)	1.91	10.1	5.30	27.9	ns
			Tc left	M2	0.75 (0.45–0.88)	7.3 (4.7–9.9)	2.14	11.0	5.93	30.5	ns
			Vc right	M1	0.99 (0.97–0.99)	4.4 (2.6–6.3)	12.00	6.9	33.30	19.1	ns
			Vc left	M1	0.99 (0.97–0.99)	5.2 (2.3–8.2)	10.93	6.7	30.30	18.7	ns
			Vc right	M2	0.97 (0.93–0.99)	6.9 (4.3–9.6)	19.00	10.7	52.70	29.8	ns
			Vc left	M2	0.98 (0.96–0.99)	7.3 (4.7–9.9)	13.17	7.9	35.51	21.4	ns
Paravlić et al., 2017	Intra- and inter-day and inter-rater reliability, n = 18				ICC, Model Yes, (95% CI)	CV	SEM	MDC	RE	Bias	
		SOL	Dm	Intra-day	0.92 (0.81–0.97)	7.5	0.63	1.80	± 0.08	ns	
				Inter-day	0.88 (0.76–95)	10.5	0.93	0.85	± 0.08	ns	
				Inter-rater	0.96 (0.90–0.98)	8.5	0.50	1.39	± 0.09	ns	
Paula Simola et al., 2016a$_{exp1}$	Inter-day reliability; 2 sessions over one week period*, n = 20				ICC, Model NM	CV	SEM	Bias			
		RF	Dm		0.92	9.3	1.0	0.811			
			Tc		0.94	4.9	1.9	0.411			

(continued on next page)

Table 10 (*continued*)

Author, Year	Type of Reliability, subjects	Muscle(s)	TMG variables	Data analysis			
				Relative Reliability	Absolute Reliability		
			Td	0.87	3.8	1.2	0.165
			V_{10}	0.92	9.0	4.0	0.544
			V_{90}	0.92	9.9	19.9	0.548
		BF	Dm	0.95	10.4	0.9	0.465
			Tc	0.91	8.7	5.6	0.089
			Td	0.92	2.4	0.8	0.635
			V_{10}	0.94	7.7	3.2	0.889
			V_{90}	0.95	7.7	11.7	0.330
		GL	Dm	0.94	13.7	0.9	0.400
			Tc	0.93	8.5	6.8	0.126
			Td	0.90	7.7	1.3	0.056
			V_{10}	0.92	12.3	3.5	0.604
			V_{90}	0.92	11.3	10.1	0.686
Rey et al., 2012	Intra-session reliability, n = 15			ICC, Model NM, (95% CI)			
		BF	Dm	0.95 (0.92–0.97)			
			Tc	0.86 (0.78–0.93)			
			Td	0.82 (0.76–0.86)			
Rodríguez-Matoso et al., 2010	Test-retest-reliability, n = 25			Cronbach's alpha			
		RF	Dm	0.920			
			Tc	0.970			
			Td	0.897			
Simunic, 2012	Between-day reliability, three consecutive days, n = 10			ICC, Model NM, (95% CI)	CV	SEM	RE	Bias
		VMO	Dm	0.98 (0.95–0.99)	4.7	± 0.17	± 0.30	ns
			Tc	0.98 (0.95–0.99)	2.2	± 0.40	± 0.56	ns
			Td	0.94 (0.82–0.98)	2.8	± 0.42	± 0.62	ns
		VL	Dm	0.99 (0.97–1.00)	4.7	± 0.25	± 0.38	ns
			Tc	0.98 (0.94–0.99)	1.5	± 0.25	± 0.41	ns
			Td	0.89 (0.69–0.97)	1.8	± 0.30	± 0.44	ns
		BF	Dm	0.99 (0.98–1.00)	4.2	± 0.43	± 0.23	ns
			Tc	0.98 (0.95–1.00)	4.9	± 1.06	± 1.50	ns
			Td	0.98 (0.87–0.99)	2.6	± 0.40	± 0.61	ns
Tous-Fajardo et al., 2010	Inter-rater and inter-electrode distance, n = 18			ICC, Model Yes, (95% CI)	CV	SEM	RE	Bias
		VM (IRR)	Dm	0.97 (0.92–0.99)	4.7	± 0.3	± 0.9	ns
			Tc	0.92 (0.81–0.97)	3.4	± 0.9	± 2.5	ns
			Td	0.86 (0.86–0.95)	2.7	± 0.9	± 2.7	ns
		VM (IED)	Dm	0.97 (0.91–0.99)	6.7	± 0.3	± 0.8	$p < 0.01$
			Tc	0.84 (0.62–0.94)	4.5	± 1.3	± 3.7	ns
			Td	0.85 (0.65–0.94)	2.0	± 0.7	± 2.0	ns
Wilson et al., 2018	Test-retest reliability, n = 10			ICC, Model NM,	CV			
		RF	Dm	0.92	5.7			

*data from submaximal stimulation are not listed. *Abbreviations:* BB, Biceps brachii; BF, Biceps femoris; CI, Confidence interval; CV, Coefficient of variation; Dm, muscle displacement; ES, Erector spinae; GM, Gastrocnemius medialis; GL, Gastrocnemius lateralis; ICC, Intra-class correlation coefficient; IED, Inter-electrode distance; IRR, Inter-rater reliability; M, time point of measurement; MDC, Minimal detectable change; NM, not mentioned; ns, not significant; N-SEM, Normalized standard error of the mean; RE, Random error; RF, Rectus femoris; SEM, Standard error of measurement; SOL, Soleus; Tc, Time of contraction; Td, Delay time; VL, Vastus lateralis; VM, Vastus medialis; VMO, Vastus medialis obliquus; Vc, V_{10} and V_{90}, Velocity of contraction.

Appell et al., 1992). Nevertheless, due to the fact that these phenomena overlap (Allen et al., 2005), and considering the limited number of studies available on this topic, we refrained from further thematic differentiation between the two types of muscle impairments.

As mentioned above, conflicting evidence was found for construct validity of the parameter Dm (Harmsen et al., 2018; Hunter et al., 2012). The researchers reported partially high post-interventional correlations (range: $r = 0.15$ to $r = 0.95$) between TMG Dm and two established biomarkers (Brancaccio et al., 2010) of muscle damage: Creatine Kinase (CK) and Myoglobin (Mb). In the review by Macgregor et al. (2018b) the authors stated that construct validity of TMG has been successfully established. By way of qualification, it must however be conceded that this statement pertains to correlations found between TMG Tc and slow-twitch muscle fibres (type I) (Dahmane et al., 2001; Dahmane et al., 2005; Simunic et al., 2011) in order to estimate muscle fiber type distribution. Nevertheless, there are further TMG publications that have assessed post-interventional significant effects on TMG measures after intense exercise (Carrasco Paez et al., 2011; Garcia-Manso et al., 2012) or in long distance sports (Garcia-Manso et al., 2011; Giovanelli et al., 2016; Gutiérrez-Vargas et al., 2018). Although the researchers did not survey data on validity, these findings substantiate our partly positive results, especially for the parameter Dm. Further research to obtain additional insights into muscles' contractile alterations as a result of exercise induced fatigue/damage is therefore

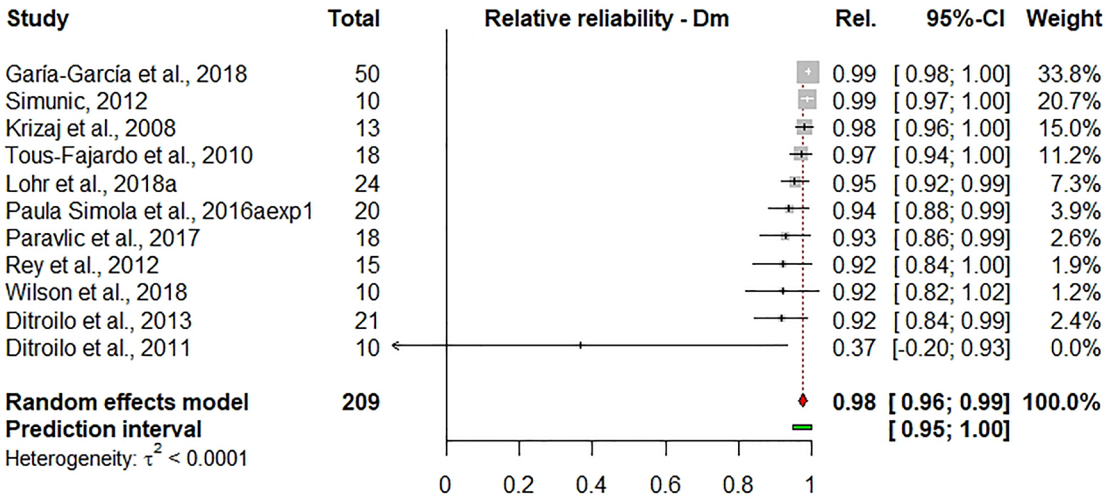

Fig. 2. Pooled inter-rater and test-retest reliabilities with 95% confidence intervals (CI) for muscle displacement (Dm).

highly desirable.

Another reason for the poor results on the instruments' ability to differentiate between fatigued and non-fatigued individuals in the three included accuracy studies as well as in the validity studies might be that the sample sizes of these studies were very limited with a range from 10 to 23 participants, which could considerably hamper the applicability of the results. Sample size determination is an important topic within the context of planning of validity and/or accuracy studies (Hajian-Tilaki, 2014). For instance, a small sample size might have limited power, and therefore generate an imprecise estimate of accuracy with wide confidence intervals (Hajian-Tilaki, 2014). The reason for the authors' choice of sample size was not provided in any of the publications on accuracy or validity. Despite the small samples, the authors of the accuracy studies concluded that TMG markers were not sensitive enough to distinguish between fatigued and recovered athletes in comparison with the criterion measure. In this regard, particular mention should be made on the study by Wiewelhove et al. (2017), a pre-post single-group study design with only 14 youth male participants. In a commentary on this publication, Martín-Rodríguez et al. (2017b) not only detected some errors within the statistical calculations, but also postulated that a larger sample size might have identified significant changes in some of the TMG measures (Dm, V_{10}, and V_{90}), and therefore, muscle fatigue could have been detected. However, based on the limited amount of studies currently available, it is not possible to give a final evaluation on this complex issue.

4.1.2. Reliability

The results of the quantitative summarization of individual reliability studies confirm the high to excellent relative reliability of all considered primary TMG measures: muscle displacement (Dm 0.98), time of contraction (Tc 0.95), and delay time (Td 0.91), demonstrating strong positive evidence. Our results regarding the relative reliability of the primary TMG variables are in line with the findings of Martín-Rodríguez et al. (2017a) who conducted the first qualitative systematic review on the instruments' reliability. They found Dm, Tc, and Td to be highly reliable, and also critically mentioned the low methodological quality of the studies included. Unfortunately, we were not able to compare the evaluated level of evidence directly, since these results were not reported by the authors. The secondary measure Vc was not included in the review by Martín-Rodríguez et al. (2017a), probably due to the limited data at the time the survey was carried out.

In contrast to relative reliability, the evidence for the primary TMG parameters' absolute reliability was less strong, especially for the spatial parameter Dm, which was rated conflicting. Possible causes for these divergent findings in terms of relative and absolute reliability will be elaborated in the further course of this section.

As previously described, the secondary TMG measure velocity of contraction (Vc, V_{10}, and V_{90}) is of particular interest as a potential measure of fatigue and recovery (Macgregor et al., 2016; Paula Simola et al., 2015; Raeder et al., 2016; Sánchez-Ureña et al., 2018). Due to the limited number of studies investigating the reproducibility of this TMG-derived parameter, and to the varying methods of calculation, we were not able to include this variable in the meta-analysis of relative

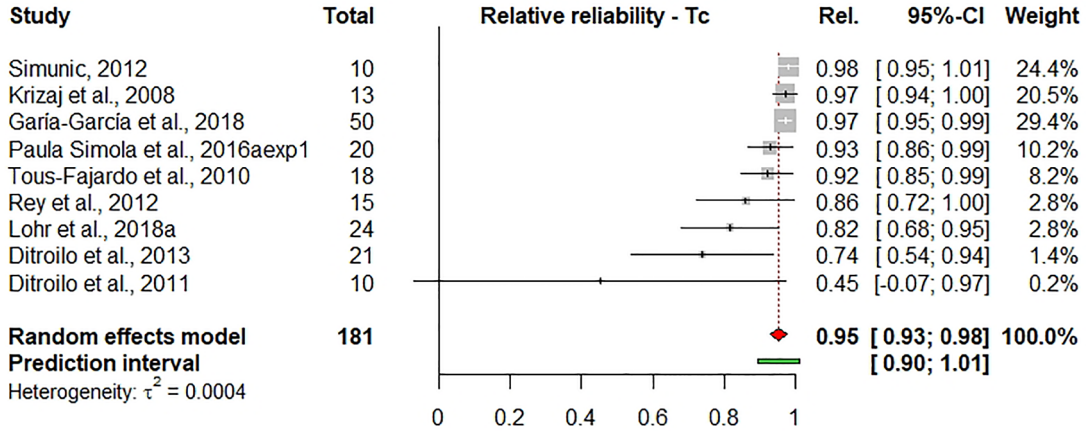

Fig. 3. Pooled inter-rater and test-retest reliabilities with 95% confidence intervals (CI) for time of contraction (Tc).

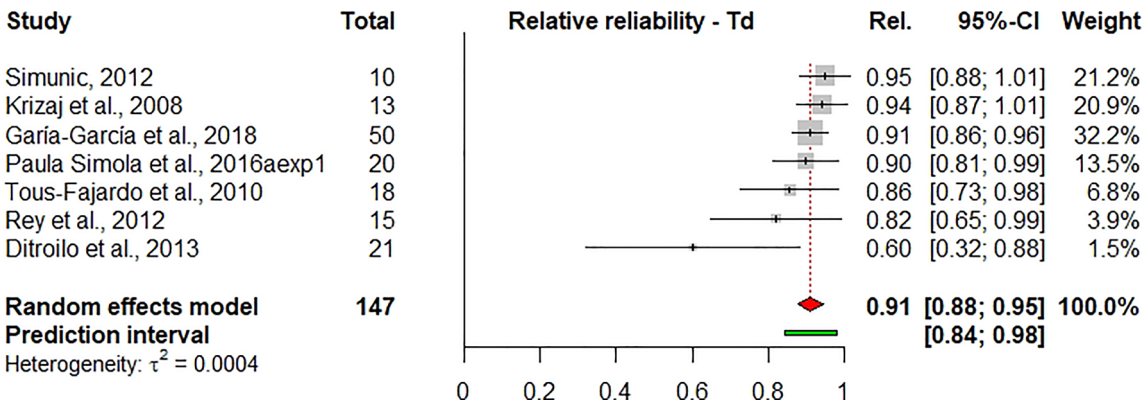

Fig. 4. Pooled inter-rater and test-retest reliabilities with 95% confidence intervals (CI) for delay time (Td).

reliability. These differences stress the need for future research to implement consistent standards to define and calculate Vc. Nevertheless, the investigation by Paula Simola et al. (2016a$_{exp1}$) and the recent reliability study conducted in our lab (Lohr et al., 2018a) offered first promising results in this regard – indicating excellent relative reliability (Vc: ICC scores ≥0.97; V_{10}: ICC scores ≥0.92; V_{90}: ICC scores ≥0.92) for this secondary measure of speed of contraction. In this regard it should be mentioned that the evidence for V_{10} and V_{90} needed to be categorized as unknown in the evidence synthesis (see Table 12) due to the low methodological quality of the study by Paula Simola et al. (2016a$_{exp1}$). Nevertheless, the results of the aforementioned investigation were entirely coherent with the variable Vc being assessed in a high quality study (Lohr et al., 2018a), and also with the relative reliability results for the primary TMG variables.

In terms of absolute reliability, the results for the parameters V_{10} and V_{90} were less unequivocal compared to Vc. In the investigation by Paula Simola et al. (2016a$_{exp1}$), V_{10} and V_{90} differed depending on the muscle under observation (cf. Tables 10 and 11). This muscle-dependent within-subject variation across different time points might be based on the muscle's specific function with varying fibre type distribution, and not least the anatomical region (Haizlip et al., 2015). Furthermore, regional specific and intra-individual subcutaneous fat distribution might have affected the measurements, with more or less pronounced within-subject differences (Wallner-Liebmann et al., 2013). These results further underline the known fact that the response to an external stimulus is individual in each skeletal muscle (Hunter, 2014).

The observed discrepancies between relative and absolute reliability for the primary and secondary TMG parameters demonstrate impressively the necessity of establishing a combination of relative and absolute indices in order to comprehensively analyse and evaluate an instruments' reproducibility (Atkinson and Nevill, 1998). The CV is a measure of within-subject variation when assessed at different time points (Atkinson and Nevill, 1998), while the ICC is indicative for the robust ability of a test or instrument to differentiate between different individuals across different time-points or raters (Weir, 2005). Based on the divergent findings for relative vs. absolute reliability, TMG can be recommended for applications where various study populations are being compared, due to its excellent relative reliability. In contrast,

Table 11
Quality of measurement properties for absolute reliability.

		Absolute reliability		Risk of bias
TMG variables	Study reference	Measurement property	rating	
Dm	Ditroilo et al., 2011	CV	−/−/−	High
	Ditroilo et al., 2013	CV	−/−/−/+	High
	García-García et al., 2018	CV	+	High
	Lohr et al., 2018a	CV	−/−/−/+	Low
	Paravlić et al., 2017	CV	+/−/+	Low
	Paula Simola et al., 2016a$_{exp1}$	CV	+/−/−	High
	Simunic, 2012	CV	+/+/+	High
	Tous-Fajardo et al., 2010	CV	+/+	Moderate
	Wilson et al., 2018	CV	+	High
Tc	Ditroilo et al., 2011	CV	−/−/−	High
	Ditroilo et al., 2013	CV	+/+/+/+	High
	García-García et al., 2018	CV	+	High
	Lohr et al., 2018a	CV	+/+/+/+	Low
	Paula Simola et al., 2016a$_{exp1}$	CV	+/+/+	High
	Simunic, 2012	CV	+/+/+	High
	Tous-Fajardo et al., 2010	CV	+/+	Moderate
Td	Ditroilo et al., 2013	CV	+/+/+/+	High
	García-García et al., 2018	CV	+	High
	Paula Simola et al., 2016a$_{exp1}$	CV	+/+/+	High
	Simunic, 2012	CV	+/+/+	High
	Tous-Fajardo et al., 2010	CV	+/+	Moderate
Vc	Lohr et al., 2018a	CV	+/+/+/+	Low
V_{10}	Paula Simola et al., 2016a$_{exp1}$	CV	+/+/−	High
V_{90}	Paula Simola et al., 2016a$_{exp1}$	CV	+/+/−	High

Abbreviations: Dm, muscle displacement; CV, coefficient of variation; Tc, contraction time; Td; delay time; Vc, V_{10}, V_{90}, velocity of contraction.

Table 12
Summary of evidence synthesis.

	Diagnostic accuracy	Validity		Reliability	
TMG variables	(*N*, F%)	Criterion validity (*N*, F%)	Construct validity (*N*, F%)	Relative (*N*, F%)	Absolute (*N*, F%)
Dm	Strong neg. (59, 34%)	Strong neg. (98, 20%)	Conflicting (29, 0%)	Strong pos.** (209, ~13%)	Conflicting*** (181, ~12%)
Tc	Strong neg. (36, 31%)	Moderate neg. (42, 26%)	Limited neg. (10, 0%)	Strong pos.** (181, ~8%)	Moderate pos.*** (153, ~9%)
Td	0	Limited neg. (20, 0%)	Limited neg. (10, 0%)	Strong pos.** (147, 9%)	Limited pos.*** (119, 11%)
Vc*	0	0	0	Moderate pos. (24, 54%)	Moderate pos.*** (24, 54%)
V_{10}*	Moderate neg. (14, 0%)	Limited neg. (14, 0%)	0	Unknown (20, 0%)	Unknown*** (20, 0%)
V_{90}*	Strong neg. (37, 24%)	Moderate neg. (37, 24%)	0	Unknown (20, 0%)	Unknown*** (20, 0%)

Levels of evidence based on van Tulder et al., 2003; *different methods of calculation; **results are based on the findings of the meta-analysis; *** results are based on the coefficient of variation; (0) not analyzed; Abbreviations: Dm, muscle displacement; F, female; *N*, total numbers of subjects; neg., negative; pos., positive; Tc, time of contraction; Td, delay time; Vc, V_{10}, V_{90}, velocity of contraction.

considering the less pronounced positive results for absolute reliability, the instrument's applicability on an individual basis, for example, for the assessment of change within single patients, cannot be approved sufficiently yet and needs to be explored in subsequent investigations.

While the results on the TMG measures' reliability are mainly good to excellent, owing to methodological constraints of a majority of the included reliability studies, there is considerable potential for bias, which has become evident from the methodological evaluation via QAREL. In the following, additional methodological issues will be highlighted. As could be shown in the present review, scientists conducting reliability investigations relating to the reproducibility of a measurement device ideally should implement a combination of relative (e.g. intra class correlation coefficients [ICC]) and absolute reliability indices (e.g. standard errors of measurement [SEM], coefficients of variation [CV]) (Atkinson and Nevill, 1998). Two of the included studies have not addressed this issue adequately, namely: Rey et al. (2012), and Rodríguez-Matoso et al. (2010). In other cases, details on the methodology of calculation (e.g. specification of the ICC model) are lacking (Paula Simola et al., 2016a$_{exp1}$; Rey et al., 2012; Wilson et al., 2018). Since different forms of ICCs can have an impact on the results when applied to an identical set of data, this potentially influences the interpretation of the results (Weir, 2005). Due to this, researchers ought to be aware of the ICC version they use and, for the sake of transparency, give adequate importance to the reporting of the 'model', 'type', 'definition', and the respective confidence intervals (Koo and Li, 2016).

4.2. Gender bias

One of the most striking observations of this systematic review was the remarkably low share of female participants (13%) in the included studies. This inconceivably low percentage of female volunteers is even considerably smaller than what is common in the current sports and exercise medicine literature, where female participants are still significantly under-represented with a share of just about 39% ($p < .000001$) (Costello et al., 2014). In fact, only four of the included studies had an even gender composition (Lohr et al., 2018a; Paravlić et al., 2017; Raeder et al., 2016; Wiewelhove et al., 2015). This affected the qualitative evaluation via QUADAS and QAREL: In case the authors refrained from explicitly defining a gender specific restriction in their eligibility criteria, or if the gender distribution was unbalanced, we consequently answered Item 1 of QUADAS/QAREL ('Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied?') with 'unclear'.

Generally it can be said that, to date, insufficient attention has been paid to sex-related differences in evidence-based (sports) medicine – this refers to study design as well as analysis (Holdcroft, 2007; Reider, 2012). Our results confirm this gender insensitivity or androcentrism, which is becoming increasingly critical considering the growing body of evidence on gender differences in health (Ruiz-Cantero et al., 2007; Verdonk et al., 2009). In sports and exercise medicine, it is well known that muscle physiological as well as pathophysiological aspects such as muscle force generation, muscle thickness, rate of muscle fatigue, and risk of injuries, inter alia, are significantly influenced by sex (Ansdell et al., 2017; Fujisawa et al., 2017; Hill et al., 2018; Hunter, 2016; Pincivero et al., 2003; Zazulak et al., 2007). This opens up the possibility of potential effects of sex-related differences on the determined TMG parameters. Due to the observed gender disparity in the included samples, there is a substantial risk of gender bias, and therefore the results and the corresponding conclusions of the studies should not be indiscriminately generalized to women, but rather be treated with caution (Holdcroft, 2007; Ruiz-Cantero et al., 2007; Verdonk et al., 2009). Investigations by (Rodríguez-Ruiz et al., 2014; 2011), and Martín-San Agustín et al. (2018) further substantiate this position (Rodríguez-Ruiz et al., 2014; 2011) evaluated muscle mechanical characteristics of the lower limbs in female and male volleyball players using TMG – the researchers found clear disparities between genders in some of the analysed muscles and TMG parameters, respectively. Recently, Martín-San Agustín et al. (2018) investigated potential sex-related differences in muscle stiffness of several lower limb muscles. They found significant ($p < .05$) dissimilarities between the female and male participants in TMG Dm of biceps femoris. Future validating research on Tensiomyography™ should therefore take these points into account by ensuring a well-balanced gender distribution during study planning and conduct. Moreover, researchers are responsible for carrying out a sex-specific analysis and interpretation of the outcomes (Ruiz-Cantero et al., 2007). Failing this, the authors ought to state clearly that the evidence from the collected data has been obtained primarily for men (Holdcroft, 2007).

4.3. Adherence to reporting standards and guidelines

In the following, several general methodological issues relating to the trial design and conduct of the incorporated studies will be discussed. There are existing reporting guidelines for the main study types in health and medical care research helping to improve the quality of reporting and publishing research. For example, the Standards for Reporting Diagnostic Accuracy (STARD) by Bossuyt et al. (2003) were developed to improve the quality of reporting on diagnostic accuracy studies. For reliability studies, eight experts in reliability and agreement investigations published the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (Kottner et al., 2011). However, only one of the studies included in this review stated the use of a guideline (Lohr et al., 2018a), although all included diagnostic accuracy and validity studies were published after the publication of STARD, and eight of the twelve reliability studies were released after the publication of GRRAS. Of course, it cannot be ruled out that the authors in fact implemented said reporting guidelines but did so without stating this adequately in their methods section. However, it is clear that the authors did not adhere to several specific reporting guidelines which will be discussed

in the following.

The assessor's qualification or competence can have a substantial impact on validity and reliability results and should therefore be transparently described (Brink and Louw, 2012; Kottner et al., 2011). In order to identify anatomical landmarks, for example, a precise and reliable palpation examination is indispensable, and in fact, this multi-layered task is not easy to learn (Browning, 2014). It could be shown that the improvement of these capabilities (i.e. reliability of examination) is facilitated by training sessions and the examiner's experience (Lavazza et al., 2018). Actually, five studies provided information in this regard, namely: Lohr et al. (2018a), Paravlić et al. (2017), Rey et al. (2012), Simunic (2012), and Tous-Fajardo et al. (2010).

Another issue refers to the insufficiently reported blinding of raters (e.g. to the reference standard). None of the included studies on diagnostic accuracy or validity has clearly documented this point. To ensure an unbiased collection of outcomes, the blinding of raters is a crucial factor (Karanicolas et al., 2010). Neglecting this may cause, inter alia, an overestimation of reliability (Lange et al., 2017).

Another procedural aspect refers to the used IED. Tous-Fajardo et al. (2010) as well as Wilson et al. (2018) demonstrated that the IED has an impact on the measurement results. Therefore, it is strongly encouraged to state the electrodes' positions very precisely – in particular, this includes the size of the electrodes and the correct measuring point (e.g. at the centre vs. the edge of the electrode). Furthermore, the researchers should document exactly whether they actually measured the IED or the distance between the electrodes and the sensor, which equals half of the IED, which in fact was not documented in one of the studies included (Harmsen et al., 2018).

A further methodological point that was scarcely reported within the studies refers to the used ISI. In case the ISI was documented, a rest period of 10 s between stimuli was chosen, with one exception: Wilson et al. (2018) used an ISI of 60 s (see Table 2). The rationale behind this procedure was to avoid the possible effect of fatigue and post-activation potentiation (Tous-Fajardo et al., 2010). According to Wilson et al. (2018), Dm and Tc can significantly be affected by the ISI. For example, Dm is significantly higher when using an ISI of 10 s compared to 30 s ($p = .017$) (Wilson et al., 2018). In order to circumvent this potential source of error, the authors support the implementation of an ISI of 30 s in upcoming investigations.

Finally, apart from the aforementioned statistical issues that should be addressed in future studies (i.e. sample size calculation, relative and absolute indices, ICC-type), it is advised to state the precision of statistical measures (e.g. measures of diagnostic accuracy or reliability) by reporting the respective confidence limits (Guyatt et al., 1995; Hess et al., 2012), because these data provide guidance for clinicians and researchers on the interpretation of a given measure. Unfortunately, confidence limits were hardly reported in the included diagnostic accuracy and validity studies (Hunter et al., 2012; Wiewelhove et al., 2017).

While not a shortcoming of any single study, we noticed one more limitation when looking at the mainly evaluated muscles as outlined in Table 2: The included studies primarily focused on limb muscles, with one exception (Lohr et al., 2018a). This is partly attributable to the easy accessibility of these anatomical regions, and moreover to the fact that muscle injury and strain of the lower limbs are very common in athletes (Zazulak et al., 2007). That is why these structures are of great interest in sport scientific research as well as for clinicians and trainers in order to support and monitor training and rehabilitation processes. Nevertheless, in the light of structural differences such as varying fibre type distributions and their respective functional requirements, different muscles respond differently in response to external stimulation (Dahmane et al., 2006) and fatiguing exercise (Hill et al., 2018). Therefore, we recommend extending further validating research to integrate and investigate a larger variety of muscular structures (e.g. trunk muscles).

Taken together, we established an overall level of evidence for the analysed TMG parameters of the studies on diagnostic accuracy, validity, and reliability of TMG to assess exercise-induced muscle fatigue by synthesizing the risk of bias and the quality of the measurement properties. The results indicate insufficient overall diagnostic accuracy for the TMG parameters Dm, Tc, V_{10}, and V_{90}. Similarly, the evidence for criterion related validity (Dm, Tc, Td, V_{10}, V_{90}) as well as construct validity (Dm, Tc, Td) was found to be either negative or inconclusive. For reliability studies, a meta-analysis was carried out; demonstrating high to excellent relative reliability for the primary TMG measures Dm, Td, and Tc. The results for the scarcely analysed secondary parameters Vc, V_{10}, and V_{90} showed evidence for good to excellent relative reliability, which, however, needs to be corroborated by additional high-quality studies, especially for V_{10} and V_{90}. In case of absolute reliability (Dm, Tc, Td, Vc, V_{10}, V_{90}) the positive evidence was less pronounced, in particular for the parameters Dm, V_{10}, and V_{90}. We evaluated a series of methodological weaknesses in the included studies. These included a striking gender bias due to the low share of female participants in the studies (Costello et al., 2014), small sample sizes (Bonett, 2002; Hajian-Tilaki, 2014), and an oftentimes poor adherence to reporting standards and guidelines (Bossuyt et al., 2003; Kottner et al., 2011).

4.4. Limitations of this review

In the current review, we have not succeeded in including exclusively high-quality studies. This is due to the limited number of studies available on the individual topics evaluated in this comprehensive systematic review. In order to evaluate the quantitative results of each measurement property we established quality criteria adapted from Mokkink et al. (2018), Prinsen et al. (2018), and Terwee et al. (2007). Originally, these quality criteria were developed for reviews of questionnaire-based patient-reported outcomes (Terwee et al., 2007). Nevertheless, numerous reviews on physical performance measures in (sport) scientific research have employed these criteria over the recent years (Drake et al., 2017; Hegedus et al., 2015; Kroman et al., 2014). Responsiveness, which is considered a further important feature of an instrument (Terwee et al., 2007), has not been addressed in the present study. In fact, exercise-induced muscle fatigue is considered as a reversible decrease of muscle force during or subsequent to work over time. Due to this temporal component in the definition of muscle fatigue, responsiveness and validity cannot conclusively be dissociated in this TMG application. A major constraint refers to the markedly small sample sizes in most of the studies included. This in fact is a common issue and major challenge in (longitudinal) investigations within sports science (Drake et al., 2017). Against this backdrop, we refrained from pre-establishing any standard about sample size requirements, as incorporated, for example, in the COSMIN guidelines (Prinsen et al., 2018). To achieve better comparability and the best possible homogeneity, the investigation was limited to healthy participants. Irrespective of our comprehensive computer-based systematic literature search, it cannot be ruled out that inadvertent sampling bias was introduced into this review, due to the fact that only one reviewer conducted the database searches, notwithstanding major efforts of the authors to assure that all relevant papers were selected. Furthermore, only one author screened the articles for inclusion, which can be considered another potential source for sampling bias. Beside this, due to limiting the search to English language articles, the possibility of language bias exists.

5. Conclusion

In summary, we conclude that robust evidence for diagnostic accuracy and criterion validity of TMG to assess exercise-induced muscle fatigue in healthy participants has yet to be established, despite several validity investigations revealing significant post-interventional correlations between the TMG variables Dm, V_{10}, and V_{90} and the reference standard of fatigue (e.g. maximal voluntary contraction). Furthermore,

our results indicate conflicting evidence for construct validity of the TMG parameter Dm, and limited negative evidence for the variables Tc and Td. The meta-analysis demonstrated that there is substantial evidence for the instruments' relative reliability; this applies to the variables Dm, Tc, and Td. With respect to absolute reliability (i.e. CV) the results for the primary TMG measures were less unequivocal, ranging from conflicting (Dm) to moderate positive (Tc). The reliability of the secondary parameters Vc, V_{10}, and V_{90} – derived from a combination of Dm and Tc or Dm, Tc, and Td – has been only occasionally examined so far. Nevertheless, based on the current findings, the few reliability studies investigating this relative measure of muscle contraction velocity indicate that Vc is highly reproducible in both relative and absolute reliability. In case of V_{10} and V_{90}, the findings available to date are not yet sufficient to prove the parameters' reliability, especially in terms of absolute reliability. Future studies implementing higher methodological standards for the conduct and reporting of TMG validation studies – including the avoidance of gender bias – are of crucial importance.

Declaration of Competing Interest

No sources of funding were used to assist in the preparation of this manuscript. The authors have no conflicts of interest that are directly relevant to the content of this manuscript.

Acknowledgements

The authors would like to thank Dr. Susann Wolff for statistical consulting and helpful comments in preparing the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jelekin.2019.05.005.

References

Alentorn-Geli, E., Álvarez-Díaz, P., Ramon, S., Marin, M., Steinbacher, G., Boffa, J.J., et al., 2015. Assessment of neuromuscular risk factors for anterior cruciate ligament injury through tensiomyography in male soccer players. Knee Surg. Sports Traumatol. Arthrosc. 23 (9), 2508–2513.
Allen, D.G., Lamb, G.D., Westerblad, H., 2008. Skeletal muscle fatigue: cellular mechanisms. Physiol. Rev. 88 (1), 287–332.
Allen, D.G., Whitehead, N.P., Yeung, E.W., 2005. Mechanisms of stretch-induced muscle damage in normal and dystrophic muscle: Role of ionic changes. J. Physiol. 567 (Pt 3), 723–735.
Andonian, B.J., Masi, A.T., Aldag, J.C., Barry, A.J., Coates, B.A., Emrich, K., et al., 2015. Greater resting lumbar extensor myofascial stiffness in younger ankylosing spondylitis patients than age-comparable healthy volunteers quantified by myotonometry. Arch. Phys. Med. Rehabil. 96 (11), 2041–2047.
Ansdell, P., Thomas, K., Howatson, G., Hunter, S., Goodall, S., 2017. Contraction intensity and sex differences in knee-extensor fatigability. J. Electromyogr. Kinesiol. 37, 68–74.
Appell, H.-J., Soares, J.M.C., Duarte, J.A.R., 1992. Exercise, muscle damage and fatigue. Sports Med. 13 (2), 108–115.
Atkinson, G., Nevill, A.M., 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 26 (4), 217–238.
Beck, T.W., Housh, T.J., Johnson, G.O., Weir, J.P., Cramer, J.T., Coburn, J.W., et al., 2006. Comparison of a piezoelectric contact sensor and an accelerometer for examining mechanomyographic amplitude and mean power frequency versus torque relationships during isokinetic and isometric muscle actions of the biceps brachii. J. Electromyogr. Kinesiol. 16 (4), 324–335.
Bonett, D.G., 2002. Sample size requirements for estimating intraclass correlations with desired precision. Statist Med 21 (9), 1331–1335.
Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Introduction to Meta-Analysis. John Wiley & Sons Ltd, Chichester, UK.
Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., et al., 2003. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Brit. M J. 326 (7379), 41–44.
Boswell, M.V., Manchikanti, L., Kaye, A.D., Bakshi, S., Gharibo, C.G., Gupta, S., et al., 2015. A best-evidence systematic appraisal of the diagnostic accuracy and utility of facet (Zygapophysial) joint injections in chronic spinal pain. Pain Phys. 18 (4), E497–E533.
Brancaccio, P., Lippi, G., Maffulli, N., 2010. Biochemical markers of muscular damage. Clin. Chem. Lab. Med. 48 (6), 757–767.
Brink, Y., Louw, Q.A., 2012. Clinical instruments: Reliability and validity critical appraisal. J. Eval. Clin. Pract. 18.
Browning, S., 2014. An investigation into the current practices and educational theories that underpin the teaching of palpation in osteopathic education: A Delphi study. Int. J. Osteopath. Med. 17 (1), 5–11.
Carlsson, H., Rasmussen-Barr, E., 2013. Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. Man Ther. 18 (2), 103–110.
Carrasco Paez, L., Sanudo Corrales, B., de Hoyo Lora, M., 2011. Tensiomyographic characteristics of rectus femoris after single bout of intense exercise. J. Soc. Sci. 7 (3), 354–357.
Cè, E., Rampichini, S., Monti, E., Venturelli, M., Limonta, E., Esposito, F., 2017. Changes in the electromechanical delay components during a fatiguing stimulation in human skeletal muscle: An EMG, MMG and force combined approach. Eur. J. Appl. Physiol. 117 (1), 95–107.
Cook, C., Cleland, J., Huijbregts, P., 2007. Creation and critique of studies of diagnostic accuracy: use of the STARD and QUADAS methodological quality assessment tools. J. Man Manip. Ther. 15 (2), 93–102.
Cook, C., Mabry, L., Reiman, M.P., Hegedus, E.J., 2012. Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. Physiotherapy 98 (2), 93–100.
Costello, J.T., Bieuzen, F., Bleakley, C.M., 2014. Where are all the female participants in Sports and Exercise Medicine research? Eur. J. Sport Sci. 14 (8), 847–851.
Dahmane, R., Djordjevic, S., Simunic, B., Valencic, V., 2005. Spatial fiber type distribution in normal human muscle Histochemical and tensiomyographical evaluation. J. Biomech. 38 (12), 2451–2459.
Dahmane, R., Djordjevic, S., Smerdu, V., 2006. Adaptive potential of human biceps femoris muscle demonstrated by histochemical, immunohistochemical and mechanomyographical methods. Med. Biol. Eng. Comput. 44 (11), 999–1006.
Dahmane, R., Valenčič, V., Knez, N., Eržen, I., 2001. Evaluation of the ability to make non-invasive estimation of muscle contractile properties on the basis of the muscle belly response. Med. Biol. Eng. Comput. 39 (1), 51–55.
Ditroilo, M., Hunter, A.M., Haslam, S., de Vito, G., 2011. The effectiveness of two novel techniques in establishing the mechanical and contractile responses of biceps femoris. Physiol. Meas. 32 (8), 1315–1326.
Ditroilo, M., Smith, I.J., Fairweather, M.M., Hunter, A.M., 2013. Long-term stability of tensiomyography measured under different muscle conditions. J. Electromyogr. Kinesiol. 23 (3), 558–563.
Doust, J.A., Pietrzak, E., Sanders, S., Glasziou, P.P., 2005. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J. Clin. Epidemiol. 58 (5), 444–449.
Drake, D., Kennedy, R., Wallace, E., 2017. The validity and responsiveness of isometric lower body multi-joint tests of muscular strength: a systematic review. Sports Med. Open 3 (1), 23.
Falco, F.J.E., Manchikanti, L., Datta, S., Sehgal, N., Geffert, S., Onyewu, O., et al., 2012. An update of the systematic assessment of the diagnostic accuracy of lumbar facet joint nerve blocks. Pain Phys. 15 (6), E869–E907.
Finsterer, J., 2012. Biomarkers of peripheral muscle fatigue during exercise. BMC Musculoskelet Disord 13, 218.
Franz, A., Behringer, M., Harmsen, J.-F., Mayer, C., Krauspe, R., Zilkens, C., et al., 2017. Ischemic preconditioning blunts muscle damage responses induced by eccentric exercise. Med. Sci. Sports Exerc. 50 (1), 109–115.
Fujisawa, C., Tamaki, A., Yamada, E., Matsuoka, H., 2017. Influence of gender on muscle fatigue during dynamic knee contractions. Phys. Ther. Res. 20 (1), 1–8.
García-García, O., Cuba-Dorado, A., Fernández-Redondo, D., López-Chicharro, J., 2018. Neuromuscular parameters predict the performance in an incremental cycling test. Int. J. Sports Med. 39, 909–915.
Garcia-Manso, J.M., Rodriguez-Matoso, D., Sarmiento, S., de Saa, Y., Vaamonde, D., Rodriguez-Ruiz, D., et al., 2012. Effect of high-load and high-volume resistance exercise on the tensiomyographic twitch response of biceps brachii. J. Electromyogr. Kinesiol. 22 (4), 612–619.
Garcia-Manso, J.M., Rodriguez-Ruiz, D., Rodriguez-Matoso, D., de Saa, Y., Sarmiento, S., Quiroga, M., 2011. Assessment of muscle fatigue after an ultra-endurance triathlon using tensiomyography (TMG). J. Sports Sci. 29 (6), 619–625.
Garrandes, F., Colson, S.S., Pensini, M., Seynnes, O., Legros, P., 2007. Neuromuscular fatigue profile in endurance-trained and power-trained athletes. Med. Sci. Sports Exerc. 39 (1), 149–158.
Gasparini, M., Sabovic, M., Gregoric, I.D., Simunic, B., Pisot, R., 2012. Increased fatigability of the gastrocnemius medialis muscle in individuals with intermittent claudication. Eur. J. Vasc. Endovasc. Surg. 44 (2), 170–176.
George, D., Mallery, P., 2003. SPSS for Windows step by step: A simple guide and reference, 11.0 update. Boston: A & B.
Giovanelli, N., Taboga, P., Rejc, E., Simunic, B., Antonutto, G., Lazzer, S., 2016. Effects of an uphill marathon on running mechanics and lower-limb muscle fatigue. Int. J. Sports Physiol. Perform. 11 (4), 522–529.
Goodwin, L.D., Leech, N.L., 2003. The meaning of validity in the new standards for educational and psychological testing. Meas. Eval. Couns Dev. 36 (3), 181–191.
Gordon, G., Holbourn, A.H.S., 1948. The sounds from single motor units in a contracting muscle. J. Physiol. 107 (4), 456–464.
Gosker, H.R., Schols, A.M.W.J., 2008. Fatigued muscles in COPD but no finishing line in sight. Eur. Respir. J. 31 (4), 693–694.
Goubert, D., de Pauw, R., Meeus, M., Willems, T., Cagnie, B., Schouppe, S., et al., 2017. Lumbar muscle structure and function in chronic versus recurrent low back pain: a cross-sectional study. Spine J. 17 (9), 1285–1296.
Grabljevec, K., Burger, H., Kersevan, K., Valencic, V., Marincek, C., 2005. Strength and endurance of knee extensors in subjects after paralytic poliomyelitis. Disabil. Rehabil. 27 (14), 791–799.
Gutiérrez-Vargas, R., Martín-Rodríguez, S., Sánchez-Ureña, B., Rodríguez-Montero, A., Salas-Cabrera, J., Gutiérrez-Vargas, J.C., et al., 2018. Biochemical and muscle mechanical postmarathon changes in hot and humid conditions. J. Strength Cond. Res.
Guyatt, G., 2008. Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice, 2nd ed. McGraw-Hill Medical, New York.

Guyatt, G., Jaeschke, R., Heddle, N., Cook, D., Shannon, H., Walter, S., 1995. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. CMAJ 152 (2), 169–173.
Haizlip, K.M., Harrison, B.C., Leinwand, L.A., 2015. Sex-based differences in skeletal muscle kinetics and fiber-type composition. Physiology (Bethesda) 30 (1), 30–39.
Hajian-Tilaki, K., 2014. Sample size estimation in diagnostic test studies of biomedical informatics. J. Biomed. Inform. 48, 193–204.
Harmsen, J.-F., Franz, A., Mayer, C., Zilkens, C., Buhren, B.A., Schrumpf, H., et al., 2018. Tensiomyography parameters and serum biomarkers after eccentric exercise of the elbow flexors. Eur. J. Appl. Physiol.
Harvill, L.M., 1991. An NCME instructional module on. Standard error of measurement. Educ. Meas. 10 (2), 33–41.
Hegedus, E.J., Cook, C., Hasselblad, V., Goode, A., McCrory, D.C., 2007. Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. J. Orthop. Sports Phys. Ther. 37 (9), 541–550.
Hegedus, E.J., McDonough, S., Bleakley, C., Cook, C.E., Baxter, G.D., 2015. Clinician-friendly lower extremity physical performance measures in athletes: A systematic review of measurement properties and correlation with injury, part 1. The tests for knee function including the hop tests. Br. J. Sports Med. 49 (10), 642–648.
Hess, A.S., Shardell, M., Johnson, J.K., Thom, K.A., Strassle, P., Netzer, G., et al., 2012. Methods and recommendations for evaluating and reporting a new diagnostic test. Eur. J. Clin. Microbiol. Infect. Dis. 31 (9), 2111–2116.
Hill, E.C., Housh, T.J., Smith, C.M., Schmidt, R.J., Johnson, G.O., 2018. Gender- and muscle-specific responses during fatiguing exercise. J. Strength Cond. Res. 32 (5), 1471–1478.
Holdcroft, A., 2007. Gender bias in research: How does it affect evidence based medicine? J. R. Soc. Med. 100 (1), 2–3.
Hunter, A.M., Galloway, S.D.R., Smith, I.J., Tallent, J., Ditroilo, M., Fairweather, M.M., et al., 2012. Assessment of eccentric exercise-induced muscle damage of the elbow flexors by tensiomyography. J. Electromyogr. Kinesiol. 22 (3), 334–341.
Hunter, J.E., Schmidt, F.L., 2006. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, 2nd ed. Sage, Thousand Oaks, Calif. [u.a.].
Hunter, S.K., 2014. Sex differences in human fatigability: mechanisms and insight to physiological responses. Acta Physiol. (Oxf) 210 (4), 768–789.
Hunter, S.K., 2016. Sex differences in fatigability of dynamic contractions. Exp. Physiol. 101 (2), 250–255.
Kallenberg, L.A.C., Schulte, E., Disselhorst-Klug, C., Hermens, H.J., 2007. Myoelectric manifestations of fatigue at low contraction levels in subjects with and without chronic pain. J. Electromyogr. Kinesiol. 17 (3), 264–274.
Karanicolas, P.J., Farrokhyar, F., Bhandari, M., 2010. Blinding: Who, what, when, why, how? Can. J. Surg. 53 (5), 345–348.
Kay, D., St Clair Gibson, A., Mitchell, M.J., Lambert, M.I., Noakes, T.D., 2000. Different neuromuscular recruitment patterns during eccentric, concentric and isometric contractions. J. Electromyogr. Kinesiol. 10 (6), 425–431.
Kersevan, K., Valenčič, V., Djordjevic, S., Simunic, B., 2002. The Muscle adaptation process as a result og pathological changes or specific training procedures. Cell. Mol. Biol. Lett. 7 (2), 367–369.
Kimberlin, C.L., Winterstein, A.G., 2008. Validity and reliability of measurement instruments used in research. Am. J. Health Syst. Pharm. 65 (23), 2276–2284.
Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15 (2), 155–163.
Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., et al., 2011. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int. J. Nurs. Stud. 48 (6), 661–671.
Krizaj, D., Simunic, B., Zagar, T., 2008. Short-term repeatability of parameters extracted from radial displacement of muscle belly. J. Electromyogr. Kinesiol. 18 (4), 645–651.
Kroman, S.L., Roos, E.M., Bennell, K.L., Hinman, R.S., Dobson, F., 2014. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: A systematic review. Osteoarth. Cartilage 22 (1), 26–39.
Lange, T., Struyf, F., Schmitt, J., Lützner, J., Kopkow, C., 2017. The reliability of physical examination tests for the clinical assessment of scapular dyskinesis in subjects with shoulder complaints: A systematic review. Phys. Ther. Sport 26, 64–89.
Langevin, H.M., Fox, J.R., Koptiuch, C., Badger, G.J., Greenan-Naumann, A.C., Bouffard, N.A., et al., 2011. Reduced thoracolumbar fascia shear strain in human chronic low back pain. BMC Musculoskelet Disord 12, 203.
Lavazza, C., Milano, V., Abenavoli, A., Maggiani, A., 2018. How type and number of training sessions influence the reliability of palpation. J. Bodyw. Mov. Ther. 22 (2), 396–401.
Leeflang, M.M.G., Deeks, J.J., Takwoingi, Y., Macaskill, P., 2013. Cochrane diagnostic test accuracy reviews. Syst. Rev. 2, 82.
Liederbach, M., Kremenic, I.J., Orishimo, K.F., Pappas, E., Hagins, M., 2014. Comparison of landing biomechanics between male and female dancers and athletes, part 2: Influence of fatigue and implications for anterior cruciate ligament injury. Am. J. Sports Med. 42 (5), 1089–1095.
Lin, C.-F., Lee, W.-C., Chen, Y.-A., Hsue, B.-J., 2016. Fatigue-induced changes in movement pattern and muscle activity during ballet releve on demi-pointe. J. Appl. Biomech. 32 (4), 350–358.
Lohr, C., Braumann, K.-M., Reer, R., Schroeder, J., Schmidt, T., 2018a. Reliability of tensiomyography and myotonometry in detecting mechanical and contractile characteristics of the lumbar erector spinae in healthy volunteers. Eur. J. Appl. Physiol. 118 (7), 1349–1359.
Lohr, C., Schmidt, T., Medina Porqueres, I., Reer, R., Braumann, K.-M., Porthun, J., 2018b. Diagnostic accuracy, Validity, and Reliability of Tensiomyography to assess muscle function and exercise-induced fatigue in healthy participants. A Systematic Review with Meta-Analysis.: PROSPERO. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=87126.
Loturco, I., Pereira, L.A., Kobal, R., Kitamura, K., Ramírez-Campillo, Zanetti V, et al., 2016. Muscle contraction velocity: a suitable approach to analyze the functional adaptations in elite soccer players. J. Sports Sci. Med. 15 (3), 483–491.
Lucas, N., Macaskill, P., Irwig, L., Moran, R., Rickards, L., Turner, R., et al., 2013. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). BMC Med. Res. Method. 13, 111.
Lucas, N.P., Macaskill, P., Irwig, L., Bogduk, N., 2010. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J. Clin. Epidemiol. 63 (8), 854–861.
Macgregor, L.J., Ditroilo, M., Smith, I.J., Fairweather, M.M., Hunter, A.M., 2016. Reduced radial displacement of the gastrocnemius medialis muscle after electrically elicited fatigue. J. Sport Rehabil. 25 (3), 241–247.
Macgregor, L.J., Fairweather, M.M., Bennett, R.M., Hunter, A.M., 2018a. The effect of foam rolling for three consecutive days on muscular efficiency and range of motion. Sports Med. Open 4 (1), 26.
Macgregor, L.J., Hunter, A.M., Orizio, C., Fairweather, M.M., Ditroilo, M., 2018b. Assessment of skeletal muscle contractile properties by radial displacement: the case for tensiomyography. Sports Med. 48 (7), 1607–1620.
Martín-Rodríguez, S., Loturco, I., Hunter, A.M., Rodríguez-Ruiz, D., Munguia-Izquierdo, D., 2017a. Reliability and measurement error of tensiomyography to assess mechanical muscle function: A systematic review. J. Strength Cond. Res. 31 (12), 3524–3536.
Martín-Rodríguez, S., Zubac, D., Piqueras-Sanchiz, F., Bautista, I.J., Simunic, B., 2017b. Commentary: tensiomyographic markers are not sensitive for monitoring muscle fatigue in elite youth athletes: a pilot study. Front. Physiol. 8, 1068.
Martín-San Agustín, R., Benítez-Martínez, J.C., Medina-Mirapeix, F., Casaña-Granell, J., 2018. Sex differences and patterns of muscle stiffness in the knee flexor and extensor musculature through analysis of isolated bellies. J. Strength Cond. Res.
Millet, G.Y., Lepers, R., 2004. Alterations of neuromuscular function after prolonged running, cycling and skiing exercises. Sports Med. 34 (2), 105–117.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2010. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int. J. Surg. 8 (5), 336–341.
Mokkink, L.B., de Vet, H.C.W., Prinsen, C.A.C., Patrick, D.L., Alonso, J., Bouter, L.M., et al., 2018. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual. Life Res. 27 (5), 1171–1179.
Neamtu, M.C., Neamtu, O.M., Marin, M.I., Bieru, D.E., 2016. Predection of motor disorders in multiple sclerosis using muscle change structure assessment. Rom. J. Morphol. Embryol. 57 (4), 1331–1335.
Orizio, C., 1993. Muscle sound: Bases for the introduction of a mechanomyographic signal in muscle studies. Crit. Rev. Biomed. Eng. 21 (3), 201–243.
Orizio, C., 2002. Comments on the letter “Accelerometer and mechanomyogram”. J. Biomech. 35 (3), 385.
Orizio, C., Gobbo, M., Diemont, B., Esposito, F., Veicsteinas, A., 2003. The surface mechanomyogram as a tool to describe the influence of fatigue on biceps brachii motor unit activation strategy. Historical basis and novel evidence. Eur. J. Appl. Physiol. 90 (3), 326–336.
Pacheco-Carrillo, A., Medina-Porqueres, I., 2016. Physical examination tests for the diagnosis of femoroacetabular impingement. A systematic review. Phys. Ther. Sport 21, 87–93.
Paravlić, A., Zubac, D., Šimunič, B., 2017. Reliability of the twitch evoked skeletal muscle electromechanical efficiency: A ratio between tensiomyogram and M-wave amplitudes. J. Electromyogr. Kinesiol. 37, 108–116.
Paula Simola, R.Á.de., Harms, N., Raeder, C., Kellmann, M., Meyer, T., Pfeiffer, M., et al., 2015. Assessment of neuromuscular function after different strength training protocols using tensiomyography. J. Strength Cond. Res. 29 (5), 1339–1348.
Paula Simola, R.Á.de., Harms, N., Raeder, C., Kellmann, M., Meyer, T., Pfeiffer, M., et al., 2016a. Tensiomyography reliability and prediction of changes in muscle force following heavy eccentric strength exercise using muscle mechanical properties. Sports Technol. 8 (1–2), 58–66.
Paula Simola, R.Á.de., Raeder, C., Wiewelhove, T., Kellmann, M., Meyer, T., Pfeiffer, M., et al., 2016b. Muscle mechanical properties of strength and endurance athletes and changes after one week of intensive training. J. Electromyogr. Kinesiol. 30, 73–80.
Peterson, K.D., Quiggle, G.T., 2016. Tensiomyographical responses to accelerometer loads in female collegiate basketball players. J. Sports Sci. 1–8.
Pincivero, D.M., Gandaio, C.B., Ito, Y., 2003. Gender-specific knee extensor torque, flexor torque, and muscle fatigue responses during maximal effort contractions. Eur. J. Appl. Physiol. 89 (2), 134–141.
Pisot, R., Narici, M.V., Simunic, B., de Boer, M., Seynnes, O., Jurdana, M., et al., 2008. Whole muscle contractile parameters and thickness loss during 35-day bed rest. Eur. J. Appl. Physiol. 104 (2), 409–414.
Pisot, R., Simunic, B., Dolenc, P., 2016. Physical inactivity effects on morphological and muskuloskeletal changes. Acta Salus Vitae 4 (1), 34–49.
Place, N., Maffiuletti, N.A., Martin, A., Lepers, R., 2007. Assessment of the reliability of central and peripheral fatigue after sustained maximal voluntary contraction of the quadriceps muscle. Muscle Nerve 35 (4), 486–495.
Portney, L., Watkins, M., 2009. Foundations of Clinical Research. Pearson Health Science, Upper Saddle River, NY.
Prinsen, C.A.C., Mokkink, L.B., Bouter, L.M., Alonso, J., Patrick, D.L., de Vet, H.C.W., et al., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual. Life Res. 27 (5), 1147–1157.
Prowse, A., Pope, R., Gerdhem, P., Abbott, A., 2016. Reliability and validity of inexpensive and easily administered anthropometric clinical evaluation methods of postural asymmetry measurement in adolescent idiopathic scoliosis: A systematic review. Eur. Spine J. 25 (2), 450–466.
Raeder, C., Wiewelhove, T., Simola, R.A.D.P., Kellmann, M., Meyer, T., Pfeiffer, M., et al., 2016. Assessment of fatigue and recovery in male and female athletes after 6 days of intensified strength training. J. Strength Cond. Res. 30 (12), 3412–3427.
Reider, B., 2012. Sex in sports medicine. Am. J. Sports Med. 40 (6), 1231–1233.
Rey, E., Lago-Penas, C., Lago-Ballesteros, J., 2012. Tensiomyography of selected lower-limb muscles in professional soccer players. J. Electromyogr. Kinesiol. 22 (6), 866–872.
Rodríguez-Matoso, D., Rodríguez-Ruiz, D., Sarmiento, S., Vaamonde, D., et al., 2010.

Reproducibility of muscle response measurements using tensiomyography in a range of positions. Rev. Andal. Med. Deporte 3 (3), 81–86.
Rodriguez-Ruiz, D., Diez-Vega, I., Rodriguez-Matoso, D., Fernandez-del-Valle, M., Sagastume, R., Molina, J.J., 2014. Analysis of the response speed of musculature of the knee in professional male and female volleyball players. Biomed Res. Int. 2014, 239708.
Rodríguez-Ruiz, D., Rodríguez-Matoso, D., Quiroga, M.E., Sarmiento, S., García-Manso, J.M., Da Silva-Grigoletto, M.E., 2011. Study of mechanical characteristics of the knee extensor and flexor musculature of volleyball players. Eur. J. Sport Sci. 12 (5), 399–407.
Rücker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M., 2008. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med. Res. Method 8, 79.
Ruiz-Cantero, M.T., Vives-Cases, C., Artazcoz, L., Delgado, A., del Mar García Calvente, M., Miqueo, C., et al., 2007. A framework to analyse gender bias in epidemiological research. J. Epidemiol. Community Health 61 (Suppl 2), ii46–ii53.
Rusu, L.D., Marin, M.N., Cosma, G.G.H., Bieru, Denisa E., Copilusi, C.P., Vasilescu, M.M., 2016. Correlation between neuromuscular assessement and evoked potential in the prediction of the multiple sclerosisi evolution. Int. J. Clin. Exp. Med. 9 (9), 17245–17257.
Sánchez-Sánchez, J., Bishop, D., García-Unanue, J., Ubago-Guisado, E., Hernando, E., López-Fernández, J., et al., 2018. Effect of a Repeated Sprint Ability test on the muscle contractile properties in elite futsal players. Sci. Rep. 8 (1), 17284.
Sánchez-Ureña, B., Rojas-Valverde, D., Gutiérrez-Vargas, R., 2018. Effectiveness of two cold water immersion protocols on neuromuscular function recovery: a tensiomyography study. Front. Physiol. 9, 766.
Schisterman, E.F., Perkins, N.J., Liu, A., Bondell, H., 2005. Optimal cut-point and its corresponding youden index to discriminate individuals using pooled blood samples. Epidemiology 16 (1), 73–81.
Seidl, L., Tosovic, D., Brown, J.M., 2017. Test-retest reliability and reproducibility of laser- versus contact-displacement sensors in mechanomyography: implications for musculoskeletal research. J. Appl. Biomech. 33 (2), 130–136.
Seijas, R., Alentorn-Geli, E., Álvarez-Díaz, P., Marin, M., Ares, O., Sallent, A., et al., 2016. Gluteus maximus impairment in femoroacetabular impingement: a tensiomyographic evaluation of a clinical fact. Arch. Orthop. Trauma Surg. 136 (6), 785–789.
Shaikh, S.A., 2011. Measures derived from a 2 × 2 table for an accuracy of a diagnostic test. J. Biom. Biostat. 2 (5).
Šimundić, A.-M., 2009. Measures of diagnostic accuracy: basic definitions. eJIFCC 19 (4), 203–211.
Simunic, B., 2012. Between-day reliability of a method for non-invasive estimation of muscle composition. J. Electromyogr. Kinesiol. 22 (4), 527–530.
Simunic, B., Degens, H., Rittweger, J., Narici, M., Mekjavic, I.B., Pisot, R., 2011. Noninvasive estimation of myosin heavy chain composition in human skeletal muscle. Med. Sci. Sports Exerc. 43 (9), 1619–1625.
Simunic, B., Krizaj, D., Narici, M., Pisot, R., 2010. Twitch parameters in transversal and longitudinal Biceps brachii response. Annales Kinesiologiae 1 (1), 61–80.
Šimunič, B., Pišot, R., Rittweger, J., Degens, H., 2018. Age-related slowing of contractile properties differs between power, endurance, and nonathletes: a tensiomyographic assessment. J. Gerontol. A Biol. Sci. Med. Sci. 73 (12), 1602–1608.
Slobogean, G.P., Verma, A., Giustini, D., Slobogean, B.L., Mulpuri, K., 2009. MEDLINE, EMBASE, and Cochrane index most primary studies but not abstracts included in orthopedic meta-analyses. J. Clin. Epidemiol. 62 (12), 1261–1267.
Stokes, M., 2009. Reliability and repeatability of methods for measuring muscle in physiotherapy. Physiother. Theory Pract. 1 (2), 71–76.
Tavakol, M., Dennick, R., 2011. Making sense of Cronbach's alpha. Int. J. Med. Educ. 2, 53–55.
Terwee, C.B., Bot, S.D.M., Boer, M.Rd., Windt, Daniëlle, A.W.M., van der, Knol, D.L., Dekker, J., et al., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60 (1), 34–42.
Than, C., Seidl, L., Tosovic, D., Brown, J.M., 2018. Test-retest reliability of laser displacement mechanomyography in paraspinal muscles while in lumbar extension or flexion. J. Electromyogr. Kinesiol. 41, 60–65.
Tosovic, D., Than, C., Brown, J.M.M., 2016. The effects of accumulated muscle fatigue on the mechanomyographic waveform: implications for injury prediction. Eur. J. Appl. Physiol. 116 (8), 1485–1494.
Tous-Fajardo, J., Moras, G., Rodríguez-Jiménez, S., Usach, R., Doutres, D.M., Maffiuletti, N.A., 2010. Inter-rater reliability of muscle contractile property measurements using non-invasive tensiomyography. J. Electromyogr. Kinesiol. 20 (4), 761–766.
Valenčič, V., Knez, N., 1997. Measuring of skeletal muscles' dynamic properties. Artif. Organs 21 (3), 240–242.
van Tulder, M., Furlan, A., Bombardier, C., Bouter, L., 2003. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine 28 (12), 1290–1299.
Verdonk, P., Benschop, Y.W.M., de Haes, Hanneke C.J.M., Lagro-Janssen, T.L.M., 2009. From gender bias to gender awareness in medical education. Adv. Health Sci. Educ. Theory Pract. 14 (1), 135–152.
Walker, S., Davis, L., Avela, J., Häkkinen, K., 2012. Neuromuscular fatigue during dynamic maximal strength and hypertrophic resistance loadings. J. Electromyogr. Kinesiol. 22 (3), 356–362.
Wallner-Liebmann, S.J., Kruschitz, R., Hübler, K., Hamlin, M.J., Schnedl, W.J., Moser, M., et al., 2013. A measure of obesity: BMI versus subcutaneous fat patterns in young athletes and nonathletes. Collegium Antropologicum 37 (2), 351–357. https://hrcak.srce.hr/file/153580.
Watakabe, M., Itoh, Y., Mita, K., Akataki, K., 1998. Technical aspects of mechnomyography recording with piezoelectric contact sensor. Med. Biol. Eng. Compu. 36 (5), 557.
Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Cond. Res. 19 (1), 231–240.
Whiting, P., Harbord, R., Kleijnen, J., 2005. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med. Res. Method. 5, 19.
Whiting, P., Rutjes, A.W.S., Reitsma, J.B., 2003. Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med. Res. Method. 3 (25), 25.
Whiting, P., Westwood, M., Beynon, R., Burke, M., Sterne, J.A., Glanville, J., 2011. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. J. Clin. Epidemiol. 64 (6), 602–607.
Wiewelhove, T., Raeder, C., de Paula Simola, Rauno Alvaro, Schneider, C., Döweling, A., Ferrauti, A., 2017. Tensiomyographic markers are not sensitive for monitoring muscle fatigue in elite youth athletes: a pilot study. Front. Physiol. 8, 406.
Wiewelhove, T., Raeder, C., Meyer, T., Kellmann, M., Pfeiffer, M., Ferrauti, A., 2015. Markers for routine assessment of fatigue and recovery in male and female team sport athletes during high-intensity interval training. PLoS ONE 10 (10), e0139801.
Wilson, H.V., Johnson, M.I., Francis, P., 2018. Repeated stimulation, inter-stimulus interval and inter-electrode distance alters muscle contractile properties as measured by Tensiomyography. PLoS ONE 13 (2), e0191965.
Wong, Y.M., 2001. Accelerometer and mechanomyogram. J. Biomech. 34 (4), 557.
Yoshitake, Y., Kawakami, Y., Kanehisa, H., Fukunaga, T., 2005. Surface mechanomyogram reflects length changes in fascicles of human skeletal muscles. Int. J. Sport Health Sci. 3 (Special_Issue_2), 280–285.
Yoshitake, Y., Ue, H., Miyazaki, M., Moritani, T., 2001. Assessment of lower-back muscle fatigue using electromyography, mechanomyography, and near-infrared spectroscopy. Eur. J. Appl. Physiol. 84 (3), 174–179.
Zarrouk, N., Chtourou, H., Rebai, H., Hammouda, O., Souissi, N., Dogui, M., et al., 2012. Time of day effects on repeated sprint ability. Int. J. Sports Med. 33 (12), 975–980.
Zazulak, B.T., Hewett, T.E., Reeves, N.P., Goldberg, B., Cholewicki, J., 2007. Deficits in neuromuscular control of the trunk predict knee injury risk: A prospective biomechanical-epidemiologic study. Am. J. Sports Med. 35 (7), 1123–1130.
Zhou, X.-H., McClish, D.K., Obuchowski, N.A., 2011. Statistical Methods in Diagnostic Medicine, 2nd ed. Wiley, op, Hoboken, N.J.

Christine Lohr is currently working on her doctoral thesis at the Department of Sports Medicine in the Faculty of Psychology and Human Movement Science at the University of Hamburg, Germany. The primary focus of her research is the application of non-invasive Tensiomyography™ (TMG) to assess and treat disorders of the myofascial system. She received her Bachelor of Science (BSc) and her Master of Science (MSc) degree in the field of Osteoopathy from the Dresden International University, Germany.

Tobias Schmidt is a qualified Physiotherapist and Osteopath. He received his medical degree (MD) from the University in Göttingen, Germany, and completed his Master's degree (M.A.) in Sports Science at the same University. He is currently working as a research fellow at the Department of General Practice, University of Göttingen, Germany. The primary focus of his research is the diagnostic and therapy of musculoskeletal disorders.

Ivan Medina-Porqueres Professor at the Department of Physical Therapy, University of Malaga, Spain. Advanced physical therapy masters in Sports Physiotherapy, Manual Therapy and Research in Sports & Physical Activity. Licensed clinical nurse and degree in Kinesiology and Physiatry. Currently working as Physical Therapist and Recovery Coach in Malaga Football Club, Malaga, Spain. His topics are Musculoskeletal Disorders Assessment, Sport Injuries Management, Myofascial Pain Approach, and Biological & Alternative Therapies.

Klaus-Michael Braumann Professor and the former Head of the Department of Sports Medicine in the Faculty of Psychology and Human Movement Science at the University of Hamburg, Germany. Currently he is the president of the German Society for Sports Medicine and Prevention (DGSP). The focus of his research is the diagnostic of overtraining, performance diagnostics and movement therapy of orthopedic and internal disorders.

Professor Rüdiger Reer, M.D., is currently head of the Department of Sports and Exercise Medicine at the University of Hamburg, Germany. He is also Secretary General of the German Association of Sports Medicine and Prevention, the largest scientific association in the field of Sports Medicine in Europe. The primary focus of his research activities comprises exercise therapy for various diseases, proprioceptive and somatosenoric abilities and neurocognition in the context of exercise.

Jan Porthun is a trained Clinical Trial Specialist, Osteopath and Medical Statistician. He received his Master of Science (MSc) in Medical Biometry from the University Heidelberg, Germany. Jan is currently an assistant professor at the Norwegian University of Science and Technology, Campus Gjøvik (NTNU-Gjøvik). The primary focus of his scientific work is the integration of innovative study designs into clinical research.