Reliability of activity monitors for physical activity assessment in patients with musculoskeletal disorders: A systematic review

Francesco Negrini\textsuperscript{a}, Alessandro de Sire\textsuperscript{b,\textasteriskcentered}, Stefano Giuseppe Lazzarini\textsuperscript{c}, Federico Pennestr\textsuperscript{a}, Salvatore Sorce\textsuperscript{d}, Chiara Arienti\textsuperscript{c} and Jacopo Antonino Vitale\textsuperscript{a}

\textsuperscript{a}IRCCS Istituto Ortopedico Galeazzi, Milan, Italy
\textsuperscript{b}Department of Medical and Surgical Sciences, University of Catanzaro “Magna Graecia”, Catanzaro, Italy
\textsuperscript{c}IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy
\textsuperscript{d}IRCCS Istituto Auxologico Italiano, Milan, Italy

Received 28 November 2020
Accepted 30 March 2021

Abstract.
BACKGROUND: Activity monitors have been introduced in the last years to objectively measure physical activity to help physicians in the management of musculoskeletal patients.
OBJECTIVE: This systematic review aimed at describing the assessment of physical activity by commercially available portable activity monitors in patients with musculoskeletal disorders.
METHODS: PubMed, Embase, PEDro, Web of Science, Scopus and CENTRAL databases were systematically searched from inception to June 11\textsuperscript{th}, 2020. We considered as eligible observational studies with: musculoskeletal patients; physical activity measured by wearable sensors based on inertial measurement units; comparisons performed with other tools; outcomes consisting of number of steps/day, activity/inactivity time, or activity counts/day.
RESULTS: Out of 595 records, after removing duplicates, title/abstract and full text screening, 10 articles were included. We noticed a wide heterogeneity in the wearable devices, that resulted to be 10 different types. Patients included suffered from rheumatoid arthritis, osteoarthritis, juvenile idiopathic arthritis, polymyalgia rheumatica, and fibromyalgia. Only 3 studies compared portable activity trackers with objective measurement tools.
CONCLUSIONS: Taken together, this systematic review showed that activity monitors might be considered as useful to assess physical activity in patients with musculoskeletal disorders, albeit, to date, the high device heterogeneity and the different algorithms still prevent their standardization.

Keywords: Fitness trackers, activity monitors, physical activity, exercise, musculoskeletal diseases
1. Introduction

Physical activity (PA) is defined as any body movement resulting in energy expenditure [1], and it is widely recognized to be of the utmost importance for the benefits it provides to physical health and general wellbeing, such as a reduced risk of developing cancer, osteoarthritis, cardiovascular disease, type 2 diabetes and depression [2,3]. Today, sedentary lifestyle is one of the main determinants of the global increase in chronic diseases, representing an individual and public health concern. Many international organizations are promoting the crucial role of PA in maintaining individual health, recommending for instance to perform 150 minutes of moderate dynamic PA per week [3,4]. In turn, more healthy people would reduce the burden of chronic disease on the sustainability of healthcare systems, cutting the rates of preventable complications and costs [5]. Then, any technology or medical device able to contribute to a significant increase in PA among patients becomes a remarkable source of shared value [6].

Gait is the cornerstone of mobility and an easy way to perform light PA. Unfortunately, several diseases (especially neurological [7,8] and musculoskeletal [9,10]) can lead to gait impairment either in quality (i.e. speed and balance) and in quantity (i.e. number of steps). Variations in gait are also influenced by psychosocial determinants, such as fear of falling, lack of a caregiver, presence of architectural barriers or other environmental factors [11]. Given the importance of walking as a function of almost any activity of day living, improving gait quality and the quantity is a fundamental healthcare outcome [12].

Although many tests have already been proposed to measure gait capacity in healthcare settings, a standard method to measure real-life performance and PA in everyday settings is still lacking, despite significant clinical benefits would follow: for instance, the possibility to measure the actual degree of PA performed by the patient at home, and where appropriate motivate him/her to increase exercise [13,14]. Therefore, identifying the most valid and reliable methods to objectively quantify the daily activities is essential. Few years ago, PA could be measured only through self-reported questionnaires, such as The International Physical Activity Questionnaire (IPAQ) [15]; the General Practice Physical Activity Questionnaire [16] or the Nordic Physical Activity Questionnaire-short [17]. However, these tools allow to collect only subjective measures of PA and they seem to overestimate the actual PA levels [18] and, further, the reliability of these tools largely depends on the reliability of patients, as PA subjective reporting may differ from data reported by objective accelerometers [15]. In light of the correlation between number of steps performed in a day and healthcare outcomes [19], an objective measure of the actual PA performed by patients is an extremely valuable information in support of clinical decision making.

In this context, the huge and recent diffusion of smartphones and smartwatches widely increased the patients’ availability of accelerometer-based PA trackers [20], even in older subjects that are typically less accustomed to technology [14]. Smartwatches are useful not only for their capability to monitor PA, but also to register acquire patient-related outcomes [21], potentially monitoring their lifestyle and health. Therefore, new technologies have been introduced in the last few years to provide an objectively measure of PA and steps per day performed by patients at home, such as accelerometers or inertial measurement units (IMU) [22], giving rise to a flourishing market of devices. Activity monitors, in particular, are portable and easy to use, and increasing studies are being published to validate them on healthy subjects [23,24] or neurological patients [14,25]; however, fewer validation studies have been published on patients affected by musculoskeletal disorders [26–28]. Musculoskeletal conditions affect, according to World Health Organization (WHO), more than 1.7 billion people and are the leading cause of disability worldwide [29]. Indeed, musculoskeletal conditions is a wide term that includes many different pathologies, more than 150 [29]; nevertheless, WHO unified them under a single term and consider one of the largest contributors to the need of rehabilitation services. Hammam et al. [30] evaluated, by means of accelerometers, PA and sedentary levels in individuals affected by rheumatoid arthritis (RA), which is one of the most common chronic inflammatory joint disease, and observed that subjects spent 9 h/day sedentary, only 0.4 h/day in moderate-to-vigorous PA and that greater sedentary time was associated with higher 10-year cardiovascular disease risk. Nevertheless, Larkin et al. [27] recently demonstrated that activity monitors significantly underestimated step counts by 26% and transition counts by 36%. In a cohort of RA patients even if the authors concluded that these instruments were valid tool for quantifying time spent in sedentary, standing or light activity, and walking behaviors in this specific population. Therefore, in light of these discrepancies, a clinical and scientifically value is mandatory for healthcare professionals, researchers, purchasers and suppli-
ers, understanding: 1) how precisely currently available commercial technologies are able to measure PA; 2) how these measures compared to subjective questionnaires in musculoskeletal patients [31]. Therefore, the aim of the present systematic review was to describe the current literature on the PA assessment through commercially available portable activity monitors in musculoskeletal patients.

2. Methods

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement [32].

2.1. Search strategy

On June 11th, 2020, PubMed, Embase, Physiotherapy Evidence Database (PEDro), Web of Science, Scopus and The Cochrane Central Register of Controlled Trials (CENTRAL) databases were systematically searched for all the articles published from inception to June 11th, 2020. The keywords that were utilized included “Musculoskeletal Diseases”, “Accelerometers”, “Fitness trackers”, “Wearable sensors”, “Exercise”, “Physical activity”, “Manual counting”, “Actigraphy”, “Steps per day”, “Sedentary”, “Moderate to vigorous”; the complete research strategy adopted is described in Table 1, adapted to each specific database.

2.2. Selection criteria

After removing duplicates, two reviewers (SS, FP) independently screened articles for eligibility. In case of disagreement, consensus was achieved by the decision of a third reviewer (JAV).

Studies were considered eligible if responding to the questions defined according to the PICOT model [33]:

1. P) The device was tested on patients affected by musculoskeletal disorders, based on the World Health Organization definition available online (https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions);
2. I) The intervention consisted in the measurement of PA by a commercially available wearable sensor based on IMUs;
3. C) The comparisons were performed against any other measurement tool (i.e. manual counting, video counting, behavioral mapping, optoelectronic system);
4. O) The outcomes measured were either number of steps per day, activity/inactivity time, or activity counts per day;
5. T) The type of study was observational.

Exclusion criteria were: 1) studies written in a language different from English; 2) full text unavailable (i.e. posters and conference abstracts).
Table 2
Reasons for exclusion of the 514 articles by the present systematic review

<table>
<thead>
<tr>
<th>Articles excluded after title and abstract screening phase (n = 498)*</th>
<th>Count (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not population of interest</td>
<td>204 (41.0)</td>
</tr>
<tr>
<td>Not intervention of interest</td>
<td>24 (4.8)</td>
</tr>
<tr>
<td>Not comparison of interest</td>
<td>432 (86.7)</td>
</tr>
<tr>
<td>Not outcome of interest</td>
<td>64 (12.9)</td>
</tr>
<tr>
<td>Not time of interest</td>
<td>218 (43.8)</td>
</tr>
<tr>
<td>Absence of abstract</td>
<td>4 (1.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Articles excluded after full-text screening phase (n = 16)</th>
<th>Count (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not population of interest</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td>Not intervention of interest</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Not comparison of interest</td>
<td>8 (50.0)</td>
</tr>
<tr>
<td>Not outcome of interest</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Not time of interest</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Absence of full text</td>
<td>7 (43.7)</td>
</tr>
</tbody>
</table>

The exclusion of the articles followed the PICOT model defined in Methods Section by the present systematic review. Data are expressed as counts (percentages). * = Papers were excluded also for more than one reason during the title and abstract screening phase.

2.3. Data extraction

Three reviewers (FN, SGL, SS) independently extracted data from included studies using a customized data extraction table in Microsoft Excel. In case of disagreement, consensus was achieved by the decision of a further reviewer (AdS).

The following data were extracted: 1) First author; 2) Publication year; 3) Location of the laboratory in which measurements took place or, if not reported, region/city where participants have been recruited; 4) Type and name of wearable device; 5) Positioning of wearable device; 6) Comparator; 7) Population and number of patients included; 8) Mean standard deviation (SD) of the age of patients; 9) Mean (SD) of their body mass index (BMI), expressed in kg/m²; 10) Inclusion criteria and, if relevant, diagnostic criteria used; 11) Validity results.

2.4. Data synthesis

Tables were used to describe both the studies’ characteristics and the extracted data. When possible, the results were gathered based on type of wearable device or population investigated.

3. Results

Out of 595 search results, 524 studies were considered eligible for inclusion after removing duplicates. Among the 524 papers screened, 498 were excluded after abstract and title screening. Kappa score of reviewer 1 and 2 was 0.168, indicating slight agreement.

Of the 26 full-text articles assessed for eligibility, 16 studies were excluded (1 had not population of interest, 8 had not comparison of interest and 8 were conference abstracts without full texts) (see Table 2 for further details). Therefore, 10 articles [27,34–41] were included in the synthesis, as depicted by PRISMA Flow Chart in Fig. 1.

All the studies have been published in the last 10 years (from 2011 to 2020). Most of them, 8 (80%), are based in Europe (4 from United Kingdom [28,35,37,41], 2 from the Netherlands [36,38], 1 from Ireland [27] and 1 from Spain [40]), 1 in America (United States of America [34]) and 1 in Oceania (New Zealand [39]).

A wide heterogeneity of wearable devices was used, taking into account that 10 different types of wearable devices were used. Nine studies investigated the role of accelerometers to assess the degree of PA in musculoskeletal patients: 5 (55.6%) studies employed different versions of the ActiGraph accelerometer (GT1M [34], GT3X [28,37], GT3X+ [40,41]); 2 (22.2%) studies employed different versions of the ActiPAL activity trackers (ActivPAL [27] and ActivPAL3 µ™ [41]); 1 study (11.1%) employed the Step-N-tune, Activ4Life and the Intelligent Device for Energy Expenditure and Activity monitors [35]; 1 study (11.1%) employed the Tri-axial accelerometer (GCdataconcepts, US) [36]; 1 study (11.1%) the Actical accelerometer [38]; and 1 study (11.1%) the Fitbit-Zip accelerometer [28]. Finally, 1 study (11.1%) assessed the number of steps per day by means of a pedometer [39].

Outcomes considered in the study included: steps per day, activity counts per day, minutes per day spent sit-
ting or performing light, moderate or vigorous activity, walking time, bouted or non-bouted physical activity, step count for 2 minutes walking test, step count during stairs test, temporal and spatial gait parameters, energy expenditure, sit to stand transfers per day.

Most of the participants suffered from rheumatic disease (RA). In particular, 6 studies [27,34,35,37,39,41] out 10 (60.0%) included patients affected by RA (total number = 464), 3 studies [34,36,39] included patients affected by osteoarthritis (OA) (total number = 237), 2 studies [35,36] included healthy subjects as controls (total number = 42). Furthermore, 3 studies included in the synthesis reported PA data from wearable devices employed on patients suffering from juvenile idiopathic arthritis [38], polymyalgia rheumatica (PMR) [28] and fibromyalgia [40].

In the first study investigating the role of wearable devices for the assessment of PA in musculoskeletal patients, Semanik et al. [34] found the data resulting from the Yale Physical Activity Scale (YPAS) to be positively correlated with the objective measurement achieved through ActiGraph GT1M accelerometer, on RA patients (r ranging from 0.00 and 0.51). Short Questionnaire to ASsess Health enhancing physical activity scores were also showed by Verlaan et al. [36] to be moderately correlated with quantitative and qualitative parameters resulting from the use of a tri-axial accelerometer (p < 0.05). Armbrust et al. [38] demonstrated that convergent validity between activity diary reports and Actical accelerometer was moderate both in case of rest and PA level (ICC = 0.41 for both). Furthermore, Larkin et al. [27] has been showed an underestimation of step counts (26%) and transition counts (36%) resulting from the use of the ActivPAL activity monitor (p < 0.001) in RA patients. Chandrasekar et al. [28] found Fitbit-Zip and GT3X+ accelerometers to systematically underestimate the number of step counts during a 2-minute-walk-test and a stairs test in PMR patients.

On the other hand, the activPAL3 µTM device accurately classified the amount of sedentary, standing and stepping time in O’Brien et al. [41], by means of...
step number (accuracy > 98%) and sit-stand transitions (accuracy = 72%). Munguía-Izquierdo et al. [28] reported an independent association of PA objective measures calculated based upon recommended vector magnitude cut point (min/week) through an ActiGraph GT3X with positive affect and life satisfaction. Lastly, Douglas-Withers et al. [39] showed that the association between PA measured by means of pedometer steps and the Health Assessment Questionnaire Disability Index score was strong (R2 = 0.562) in RA patients, while association (R2 = 0.156) was characteristic of OA patients.

Further details on wearable devices type and positioning, outcome measures assessed, comparator, mean age and BMI of study population, inclusion criteria, and validity results of all studies included by the present systematic review are depicted in Supplementary Table 1.

4. Discussion

The present systematic review aimed to describe the commercially available portable activity trackers currently employed to measure the levels of PA in musculoskeletal patients. All the studies included in the synthesis were published in the last 10 years, confirming how these technologies are gaining momentum. Being their validity still under scrutiny, the authors searched the literature for papers reporting a comparison with other instruments. Only 3 studies compared portable activity trackers with objective measurement tools such as the GAITrite instrumented walkway [35] or manually counted steps [27,35,41]. All the other 7 studies reported comparisons with subjective questionnaires regarding PA, Quality of life or disability [28,34,36–40].

A high percentage of studies (6 out of 10, 60.0%) were related to RA patients [27,34,35,37,39,41], probably because moderate PA can be beneficial for their health (i.e., improving endothelial function, slowing atherosclerotic process, slowing radiographic progression and increasing mineral bone density) [42]. The study by Larkin et al. [27] underlines how the activPAL activity monitor – although underestimating the number of step and transition counts, which is peculiar to this type of device [14,27,28] – is valid to measure the amount of time spent in sedentary, standing, and walking behaviors in people with RA [27,41]. Conversely, YPAS questionnaire that subjectively assess PA in RA showed inconsistent correlations with the levels of PA estimated by objective devices [34,37]. However, this result can be explained by the different impact of the same disease on the functional capabilities of patients [43]. The lack of differentiation between different degrees of functional impairment in RA patients is common to all the papers included in the synthesis, calling for further research in order to establish comparable gait patterns.

The review clearly shows most of the limitations that prevent from implementing the use of activity trackers during every day clinical practice. The first and foremost issue is the lack of instrument standardization, both from a hardware and a software standpoint. Many different types of activity tracker have been proposed, with different positioning and different algorithms employed to support the interpretation of data. These algorithms are never clearly described in any paper, although this is fundamental in order to obtain a standardized protocol for the objective measurement of PA. As long as each device introduced in clinical practice needs a specific validation, the process of evidence-based comparison and practice will be severely slowed down. The different algorithms utilized in the studies included cause the different outcomes analyzed in the paper: steps per day, activity counts per day, minutes per day spent sitting or performing light, moderate or vigorous activity, walking time, bouted or non-bouted physical activity [44] are all used in different combination in the included papers. Furthermore, the inclusion criteria adopted within the studies do not clearly define either the gait pattern or velocity of patients, which is known to vary significantly among different patients and/or stages of the same disease.

Activity monitors are mostly based on IMUs which need to work in combination with a dedicated software to analyze raw data [22]. However, it has been shown that different gait patterns and especially different velocity can prevent the correct detection of PA levels [14]. For this reason, future studies assessing the validity of activity tracker should focus on the universal functional criteria represented by the International Classification of Functioning, Disability and Health [45], rather than criteria derived from each single disease.

On the other hand, all the studies included in the synthesis have considered the number of steps or counts per day as a common quantitative outcome, and the number of minutes of sedentary time, light and moderate or vigorous PA per day spent sitting or performing light, moderate or vigorous activity, walking time, bouted or non-bouted physical activity [44] are all used in different combination in the included papers. Furthermore, the inclusion criteria adopted within the studies do not clearly define either the gait pattern or velocity of patients, which is known to vary significantly among different patients and/or stages of the same disease.

The review clearly shows most of the limitations that prevent from implementing the use of activity trackers during every day clinical practice. The first and foremost issue is the lack of instrument standardization, both from a hardware and a software standpoint. Many different types of activity tracker have been proposed, with different positioning and different algorithms employed to support the interpretation of data. These algorithms are never clearly described in any paper, although this is fundamental in order to obtain a standardized protocol for the objective measurement of PA. As long as each device introduced in clinical practice needs a specific validation, the process of evidence-based comparison and practice will be severely slowed down. The different algorithms utilized in the studies included cause the different outcomes analyzed in the paper: steps per day, activity counts per day, minutes per day spent sitting or performing light, moderate or vigorous activity, walking time, bouted or non-bouted physical activity [44] are all used in different combination in the included papers. Furthermore, the inclusion criteria adopted within the studies do not clearly define either the gait pattern or velocity of patients, which is known to vary significantly among different patients and/or stages of the same disease.

Activity monitors are mostly based on IMUs which need to work in combination with a dedicated software to analyze raw data [22]. However, it has been shown that different gait patterns and especially different velocity can prevent the correct detection of PA levels [14]. For this reason, future studies assessing the validity of activity tracker should focus on the universal functional criteria represented by the International Classification of Functioning, Disability and Health [45], rather than criteria derived from each single disease.
performance in patients with musculoskeletal disorders [46,47].

A 2015 systematic review on reliability of activity monitors on healthy subject showed a high inter-device reliability for some of the outcomes most commonly studied with activity monitors such as steps, distance, energy expenditure, and sleep [48]. However, in the present paper we aimed to study patients suffering from musculoskeletal diseases. While healthy subjects have usually a physiological gait pattern, patients have often pathological gait pattern characterized by asymmetry, thus harder to detect by activity monitors.

Therefore, findings of our study are aligned with a recent Cochrane Systematic Review [49] focused on activity monitors as instruments for increasing physical activity in adult stroke survivors. Linch et al. underline the role of activity monitors in stroke patients to increase their PA, with the obvious benefit on health-related outcomes. They concluded that, despite activity monitors are promising tools to increase PA in stroke patients, still could not be recommended in everyday clinical practice due to heterogeneity and limitations of studies on the topic. It has to be underlined that we detected the aforementioned heterogeneity in scientific papers regarding activity monitors, albeit study aim and population are different from their systematic review [49]. Ultimately, our study describes the state-of-the-art of findings in PA monitoring patients suffering from musculoskeletal disorders, the different commercially available activity monitors already tested, and the new outcomes they are able to provide to clinicians and patients, that could be eventually implemented in the everyday clinical practice.

4.1. Study limitations

Limitations of the study are linked to the limited number of papers found that met inclusion criteria and their heterogeneity that prevent us from being able to perform a meta-analysis. Furthermore, risk of bias was not performed because of the lack of specific scales for the type of study included.

5. Conclusions

Taken together, findings from papers included in the present systematic review suggested that activity monitors might be considered as viable methods to assess PA in patients affected by musculoskeletal disorders. Clinicians currently have at their disposal many different commercially available activity monitors, and a few of them have already been tested on actual patients as showed in the present systematic review. Our review can inform physicians and patients about results and outcomes of utilizing activity monitors on musculoskeletal disorders. However, a further methodological standardization is still needed in order to accelerate the process of evidence-based comparisons in support of the common clinical practice.

Author contributions

Jacopo Antonino Vitale and Chiara Arienti equally contributed as last authors.

Conflict of interest

None of the authors declare any conflict of interests for the research, authorship, and/or publication of this article.

Funding

None of the authors have received any financial support for the research, authorship, and/or publication of this article.

Supplementary data

Supplementary Table 1 is available from https://dx.doi.org/10.3233/BMR-200348.

References


