



OPEN ACCESS

EDITED BY

Behdin Nowrouzi-Kia,
University of Toronto, Canada

REVIEWED BY

Ruopeng Sun,
Stanford University, United States
Zhuoying QIU,
China Rehabilitation Research Center/WHO
Collaborating Center for Family International
Classifications, China

*CORRESPONDENCE

Ashley Polhemus
ashley.polhemus@uzh.ch

SPECIALTY SECTION

This article was submitted to Translational
Research in Rehabilitation, a section of the
journal Frontiers in Rehabilitation Sciences

RECEIVED 08 July 2022

ACCEPTED 01 November 2022

PUBLISHED 28 November 2022

CITATION

Polhemus A, Haag C, Sieber C, Sylvester R,
Kool J, Gonzenbach R and von Wyl V (2022)
Methodological heterogeneity biases physical
activity metrics derived from the Actigraph
GT3X in multiple sclerosis: A rapid review and
comparative study.
Front. Rehabil. Sci. 3:989658.
doi: 10.3389/fre.2022.989658

COPYRIGHT

© 2022 Polhemus, Haag, Sieber, Sylvester, Kool,
Gonzenbach and von Wyl. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Methodological heterogeneity biases physical activity metrics derived from the Actigraph GT3X in multiple sclerosis: A rapid review and comparative study

Ashley Polhemus^{1*}, Christina Haag^{1,2}, Chloé Sieber^{1,2},
Ramona Sylvester³, Jan Kool³, Roman Gonzenbach³
and Viktor von Wyl^{1,2}

¹Epidemiology and Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland,

²Institute for Implementation Science in Health Care, University of Zürich, Zürich, Switzerland,

³Research Department Physiotherapy, Rehabilitation Centre, Valens, Switzerland

Background: Physical activity (PA) is reduced in persons with multiple sclerosis (MS), though it is known to aid in symptom and fatigue management. Methods for measuring PA are diverse and the impact of this heterogeneity on study outcomes is unclear. We aimed to clarify this impact by comparing common methods for deriving PA metrics in MS populations.

Methods: First, a rapid review of existing literature identified methods for calculating PA in studies which used the Actigraph GT3X in populations with MS. We then compared methods in a prospective study on 42 persons with MS [EDSS 4.5 (3.5–6)] during a voluntary course of inpatient neurorehabilitation. Mixed-effects linear regression identified methodological factors which influenced PA measurements. Non-parametric hypothesis tests, correlations, and agreement statistics assessed overall and pairwise differences between methods.

Results: In the rapid review, searches identified 421 unique records. Sixty-nine records representing 51 eligible studies exhibited substantial heterogeneity in methodology and reporting practices. In a subsequent comparative study, multiple methods for deriving six PA metrics (step count, activity counts, total time in PA, sedentary time, time in light PA, time in moderate to vigorous PA), were identified and directly compared. All metrics were sensitive to methodological factors such as the selected preprocessing filter, data source (vertical vs. vector magnitude counts), and cutpoint. Additionally, sedentary time was sensitive to wear time definitions. Pairwise correlation and agreement between methods varied from weak (minimum correlation: 0.15, minimum agreement: 0.03) to perfect (maximum correlation: 1.00, maximum agreement: 1.00). Methodological factors biased both point estimates of PA and correlations between PA and clinical assessments.

Conclusions: Methodological heterogeneity of existing literature is high, and this heterogeneity may confound studies which use the Actigraph GT3X. Step counts were highly sensitive to the filter used to process raw accelerometer data. Sedentary time was particularly sensitive to methodology, and we recommend using total time in PA instead. Several, though not all, methods for deriving light PA and moderate to vigorous PA

yielded nearly identical results. PA metrics based on vertical axis counts tended to outperform those based on vector magnitude counts. Additional research is needed to establish the relative validity of existing methods.

KEYWORDS

actigraph, physical activity, step count, methodology, multiple sclerosis

1 Introduction

Multiple sclerosis (MS) is a neurodegenerative autoimmune disease which affects individuals' physical and cognitive function, motor control, and energy levels. Physical activity (PA) is known to aid in symptom management and improve quality of life in individuals with MS (1, 2). However, PA is reduced in this population compared to healthy controls (3, 4). Interventions which can increase PA in individuals with MS, for example through rehabilitation (5), exercise training (2), or behavior modification coaching (6), are therefore a topic of public health and research interest (6). However, if we are to understand PA behavior in MS populations and to evaluate the efficacy of these interventions, we require PA outcomes measures which are valid, responsive, and feasible to implement.

Many outcome measures have been used to measure PA in populations with MS (7, 8), though the optimal PA metrics for MS populations are not yet clear (6). Patient reported measures, such as the International Physical Activity Questionnaire (IPAQ) (9) or Godin's Leisure Time Exercise Questionnaire (GLTEQ) (10), are inexpensive and validated in MS, but are burdensome to complete regularly, subject to recall bias, and insensitive to short bouts of light or lifestyle PA (11–14). Further, these measures assess participants' perceptions of their own PA behavior, rather than the PA they objectively completed. Alternatively, objective metrics derived from wearable sensors can quantify PA unobtrusively during daily life with minimal input from the wearer. The Actigraph GT3X is the most common sensor used for PA research with persons with MS (PwMS) (7). Unlike most other PA trackers, the Actigraph is capable of deriving step counts, PA intensity, and other metrics in a flexible manner, allowing researchers substantial freedom in the way they process data and calculate PA.

While this flexibility allows the Actigraph to be used in many settings and patient populations, it also contributes to substantial methodological heterogeneity in the PA literature. In a systematic review, Migueles et al. mapped the sensor settings and data processing methods used to derive PA metrics from the Actigraph for healthy populations (15). These differences included various sensor placements, data filtering methods, wear time definitions, and cutpoints (i.e., thresholds which differentiate sedentary behavior, light PA [LPA], and moderate to vigorous PA [MVPA]). These

methodological factors have the potential to affect the outcomes of Actigraph-derived PA metrics at various points throughout study conduct and data analysis (Table 1). In the MS literature, the extent of this methodological heterogeneity and its practical implications, both for study design and the interpretation of existing literature, are currently unclear.

The objective of this study was to quantify the extent and the potential confounding effects of methodological heterogeneity on PA metrics derived from the Actigraph GT3X in MS populations. First, we conducted a rapid review to assess the methodological heterogeneity of the literature and to identify common data processing methods. We then directly compared these data processing methods by applying them on a single dataset, quantifying the impact of heterogeneity on six PA metrics. These analyses identified biases arising from methodological heterogeneity and suggest implications for the implementation, interpretation, and meta-analysis of PA metrics in MS populations.

2 Methods

2.1 Rapid review

The purpose of this review was to rapidly identify established data processing methods and their common derivatives for calculating PA metrics with the Actigraph GT3X in PwMS. It was modeled after the systematic review conducted by Migueles et al. (15), and adhered to rapid review conduct guidelines (22, 23). In July 2020 and again in September 2021, we searched PubMed, Embase, IEEE Xplore for peer-reviewed and gray literature published in the English language during or after 2010, reflecting the release of the Actigraph GT3X in mid-2009 (15) and its validation for MS populations in 2012 (24, 25). We supplemented this corpus with the first 100 hits from Google Scholar and with manual reference searches. All search strategies are provided in the supplementary materials. Records were eligible if they reported using the Actigraph GT3X to measure PA in a population which included adult PwMS in the English language. Any study design, including research protocols and validation studies, were eligible. Eligibility did not depend on study setting, PA metrics, diagnostic criteria, disease severity, or disease duration. Records were screened and managed in DistillerSR software.

TABLE 1 Methodological factors which can affect PA metrics derived from the Actigraph GT3X during device initialization, study conduct, and data processing.

Implementation phase	Methodological factors	Description
Device Initialization	Sampling frequency	The Actigraph measures acceleration a set number of times per second (i.e., the sampling frequency), generating raw data in three axes, which correspond to movement along the body's longitudinal, sagittal, and frontal axes. The sampling frequency is selected before the device is given to the participant (16).
Study Conduct	Sensor placement	The Actigraph can be worn on either hip, wrist, or ankle. The locations do not produce equivalent data, and data processing methods are generally validated with respect to a specific sensor placement (17, 18).
Data Processing		
Filter data	Filter	During processing, a bandpass filter is applied to the raw data to remove non-physiological movement artifacts. Two settings are available: a standard filter for healthy gait and the low-frequency extension (LFE) for populations with impaired gait. The LFE is designed to be more sensitive to light, slow movements (16).
Aggregate into epochs	Epoch length	Filtered data is aggregated into epochs, which can be seconds, minutes, or longer in length. The resulting data is expressed in terms of the number of activity counts or steps which occurred during each epoch. Many algorithms are designed to be used with specific epoch lengths (16).
Identify non-wear time	Wear time algorithm Inactivity period	Data is then further processed through wear time algorithms to differentiate between inactive periods when the wearer was sedentary, and those when the device was not worn. This typically consists of identifying extended periods with zero activity counts, with or without a spike tolerance (19).
Remove invalid days	Minimum daily wear time	A minimum daily wear time is typically specified to ensure that PA metrics are valid representations of the wearer's daily activity. Days which did not reach this minimum are usually considered invalid and removed from the dataset (16).
Remove invalid measurements	Minimum valid days Minimum weekdays vs. weekend days	Daily PA metrics are frequently averaged over the course of a week to account for day-to-day variability. A minimum number of wear days is typically specified to ensure measures are not confounded by this variability. Studies sometimes require that measurements do or do not contain a weekend day, as activity can differ from weekdays (20).
Calculate physical activity metrics	Data source Algorithms, cutpoints, and metric definitions	Algorithms and cutpoints (i.e., limits used to differentiate between PA intensity levels) are applied to the data to calculate PA metrics. These methods may use raw or epoch-level data, vertical axis activity counts, vector magnitude counts, step counts, or other data sources to calculate PA metrics. Further aggregation and processing are then applied to calculate outcome measures, such as average daily metrics (16). Researchers may opt to adjust PA metrics for wear time if wear time is contributing to intra- and inter-participant variability (21).

A single reviewer screened all abstracts and full texts for eligibility, critically appraised records, and extracted data. In contrast to systematic reviews, single-screening may be used for rapid reviews (22, 23). To our knowledge, no appraisal tool exists for critically appraising technology implementation methods. Therefore, we assessed quality through two questions:

1. Were sensor-related methods sufficiently clear to be reproducible?
2. Did the study adhere to previously validated methods? If not, did they provide novel evidence of validity?

The same single reviewer extracted data from each eligible study, including a brief description of the study design, aims, and findings, characteristics of the study population, device placement, sampling frequency, filter settings, non-wear time definition, derived PA metrics, cutpoints used for activity intensity classification, and other algorithms used to calculate PA measurements. Multiple records arising from the same study were reported separately, as they frequently reported different data processing and analysis methods. The

relationships between records arising from the same study are noted in the supplementary materials. Methodological heterogeneity was evaluated qualitatively through frequency analysis (i.e., number of studies which used a given method) and narrative analysis. Frequency analysis was conducted both overall and per PA metric to identify whether methodological heterogeneity was consistent across metrics. No further sensitivity analyses or risk of bias analyses were conducted.

The protocol was finalized prior to study conduct, but was not registered as systematic review repositories do not accept rapid review protocols. No amendments were made after the protocol was finalized.

Common data processing methods, or those repeatedly used in the literature, were identified based on frequency analysis and the references in included studies. For the purposes of this study, we refer to a "method" as the unique combination of methodological factors (i.e., filter, data source, cutpoint, wear time definition, etc.) that were used to calculate each point estimate of each PA "metric" (i.e., step count, time in PA, sedentary time, etc.).

2.2 Experimental study

2.2.1 Study participants and procedures

This study is a sub-study of BarKA-MS, a cohort study exploring barriers to PA in PwMS during and after neurorehabilitation (26). We recruited a convenience sample of PwMS undergoing elective inpatient neurorehabilitation at the Kliniken Valens between January and November 2021. Participants were eligible if they (1) had a clinically confirmed diagnosis of MS, (2) were 18 years of age or older, (3) had reduced walking ability but were able to walk independently with or without an assistive device, (4) had access to WiFi and a mobile device in the rehabilitation center and at home, (5) were willing to wear study devices to measure their PA, and (6) were able to answer study questionnaires in German. The ethics committee of the canton of Zurich approved the study protocol (BASEC-no. 2020-02350) and all participants provided written informed consent.

Screening, enrollment, and baseline visits were conducted at the rehabilitation clinic. Participants then completed personalized rehabilitation programs, which lasted between two and four weeks. For the purposes of this analysis, we collected the following outcome measures at the end of participants' rehabilitation stay: Extended disability status scale (EDSS) score (27), the 10 meter gait speed test (10 mWT) (28), the 6 min walk test (6 MWT) (29), the Multiple Sclerosis Walking Scale – 12 (MSWS-12) (30), a measure of walking ability and its impact on daily activities, and the International Physical Activity Questionnaire (IPAQ) (9), a self-assessment of PA in the previous seven days which was previously validated in PwMS.

Approximately one week prior to discharge, participants were fitted with an Actigraph GT3X accelerometer (Manufacturing Technology, Inc., FL, USA), which was worn on the non-dominant hip for the last week of the rehabilitation stay. The day the device was provisioned and the day participants were discharged were not included in the analysis, leaving most participants with a potential measurement duration of five to seven days. Participants occasionally wore the devices for a longer period if their rehabilitation was unexpectedly extended. Devices were initialized in Actilife Version 6.0 at a sampling rate of 30 Hz.

We opted to use data collected during rehabilitation, rather than in the home environment, in our primary analyses because participants were expected to engage in a diverse set of structured and unstructured physical activities during their rehabilitation programs. This choice allowed us to study PA metrics on a dataset which was known to capture sedentary behavior, LPA, and MVPA, with high wear compliance. Data collected in the home exhibited lower wear compliance and fewer valid days, with additional participants not meeting minimum wear requirements. However, this also means that the average PA metrics presented here are not necessarily

representative of participants' normal behavior, and should not be interpreted as such.

2.2.2 Data processing

Data were uploaded to Actilife, processed *via* two filters [the standard filter and the low frequency extension (LFE)] and aggregated into 60 s epochs. Actigraph's proprietary step detection algorithm counted steps per minute on data derived from each filter. Data were then exported for processing in R (Version 4.1.0). Individual epochs were then categorized as sedentary, LPA, or MVPA according to the methods identified in the review (see review results, Table 3). This process yielded multiple sets of data for each PA metric (Step count: 2 sets, activity counts: 4 sets, total time in PA: 12 sets, sedentary time: 12 sets, time in LPA: 18 sets, time in MVPA: 8 sets).

Though we tested the impact of wear time definitions on PA metrics, our primary definition of non-wear time was 60 min of continuous zeros. Days with at least 10 h of wear time during waking hours (6AM to 11PM) were considered valid and a minimum of two valid days was required for inclusion (31, 32). We did not require measurements to include weekend days. This definition was used to determine whether participants met minimum wear time requirements, and was the default wear time definition used in analyses which did not assess the effects of wear time definition. We then calculated wear time *via* each combination of wear time algorithms and inactive periods ($n=10$) for each data set derived for each PA metric. Epochs during which the Actigraph was not worn were removed. This process resulted in a unique data set which represented each method, or unique combination of methodological factors (Step count: 20 sets, activity counts: 40 sets, total time in PA: 120 sets, sedentary time: 120 sets, time in LPA: 180 sets, time in MVPA: 80 sets). For each data set, each participant's epoch-level data were then aggregated into daily PA metrics by adding all epoch-level PA during valid wear time over the course of waking hours each day. Daily metrics were then averaged over all valid days to generate average PA metrics. Average PA metrics were used in all analyses.

In some cases, researchers adjust PA metrics for wear time by expressing PA either as ratio (i.e., PA per unit wear time) which can be scaled to reflect standardized daily PA metrics to account for unequal wear times across days and across participants, reducing variability at the participant and population level (21). However, if methodological heterogeneity induces bias in wear time estimates, this method has the potential to further bias PA metrics. To quantify these effects, we generated a separate dataset by dividing each daily PA metric by the wear time measured on that day according to each of the possible wear time estimates. To allow for comparability across adjusted and non-adjusted PA metrics, we assumed a standard daily wear

length of 13 h and multiplied the PA ratios by this standardized day length. We chose 13 h as a “standard” day length because this was approximately the average daily wear time of the Actigraph in our study. We adopted this approach, rather than using a proportion of wear time, to enable better comparability across adjusted and unadjusted PA metrics.

2.2.3 Statistical analysis

The aims of the statistical analyses were twofold: (1) to identify the methodological factors which independently influenced PA estimates and (2) to understand the potential confounding effects of methodological heterogeneity. All statistical analyses were conducted in R (Version 4.1.0) using the RStudio environment (Version 1.4.1717).

First, multivariate linear regression with participant-level random intercepts was used to identify methodological factors which independently influenced PA metrics. For each PA metric, all replicate data sets were combined and the methodological factors used to derive these sets were treated as independent variables. Data were derived from the same raw accelerometer data and therefore highly inter-dependent, making models at risk for over-fitting. Therefore model development was based on theoretical knowledge and Bayesian Information Criterion (BIC), rather than p values. A base model tested the effects wear time algorithm, inactivity period definitions, data source, filter, and cutpoints, as relevant for each metric (see review results, [Table 3](#)). Each base model was adjusted for the participants’ MS severity, defined as mild (EDSS < 4.0), moderate (EDSS 4.0–5.5), and severe (EDSS 6.0–6.5) body function impairment, consistent with previous studies ([33](#)). Interactions between methodological factors were considered plausible, as was effect modification by MS severity. We therefore tested for interactions between model terms through a manual forward stepwise procedure. Interactions between methodological factors were tested first, followed by interactions between methodological factors and disease severity strata. Interaction terms were retained if the more complex model exhibited a lower BIC than the previous, simpler model. We refer to models containing all methodological factors and eligible interaction terms as “fully adjusted” models. When methodological factors did not contribute substantially to these models, (i.e., when coefficients were smaller than 5% of the intercept and the factor did not exhibit interactions), these terms were removed from the model in a backward stepwise manner according to BIC. When terms did not contribute substantially to the fully adjusted model, the process was repeated in a reduced dataset which excluded all combinations of methods with these variables. The resulting models are referred to as “reduced” models. Reduced models were considered the primary models for this analysis, but both sets of models are reported in the supplementary material. This conservative model development process reduced the risk of

over-fitting and spurious findings related to the close relationship between replicate data sets. All p -values are reported, and were corrected for multiple testing through a Benjamini Hochberg procedure ([34](#)). The R package `lmerTest` was used to build these models. This process was repeated for PA metrics which were adjusted for wear time.

Subsequent analyses were limited to combinations of methodological factors which were included in unadjusted PA metrics’ reduced models. For each PA metric, Friedman tests identified the presence of significant differences between methods. Subsequent pairwise analyses identified median differences, statistically significant differences (Wilcoxon tests), correlations (Pearson’s correlation), and agreement [Lin’s concordance correlation coefficient ([35](#)) (CCC)] between each pair of methods. With the exception of sedentary time estimates, the effects of wear time algorithm and inactivity period were not included in this analysis, as they did not substantially affect PA estimates (see results). All p -values were corrected for multiple testing through a Benjamini Hochberg procedure ([34](#)). All pairwise analyses used the R package `stats`, except for the CCC analyses which used the package `DescTools`.

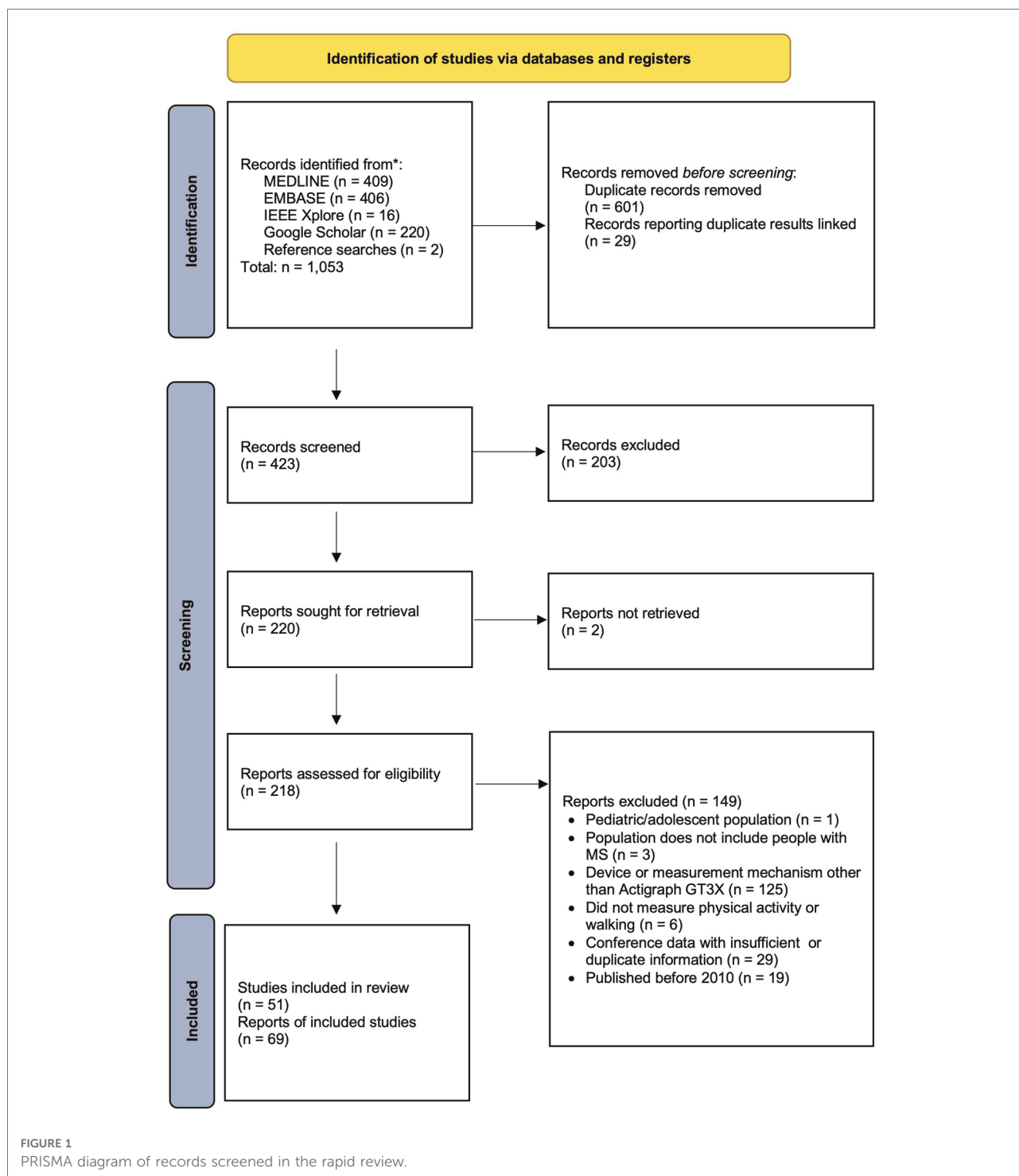
To illustrate the potential confounding effects of methodological heterogeneity, we conducted several analyses which are relevant to recent MS literature and policy development. First, we explored the impact of methodological heterogeneity on the associations between PA metrics and five clinical outcome measures (IPAQ, EDSS, 10 meter gait speed, 6MWT, and MSWS-12) which are known to associate with PA. We calculated the Spearman correlation between these clinical measures and all PA metrics and methods. We then made qualitative comparisons between point estimates of the correlations across methods for each metric. In addition, several studies have investigated whether people with MS meet minimum PA recommendations ([36, 37](#)). The National MS Society recommends that ambulatory PwMS take 7,500 steps per day or engage in 150 min of MVPA per week (i.e., 30 min per day 5 times per week) ([3](#)). We therefore compared the proportion of the study population which met these guidelines according to each method for deriving step count and MVPA. The other metrics assessed here do not have established guidelines, and were not assessed in this analysis.

3 Results

3.1 Rapid review

3.1.1 Assessing methodological heterogeneity in the literature

Searches identified 421 unique records, of which 216 were included in full text review. From these, 69 records from 51 unique studies were eligible ([Figure 1](#)) ([20, 24, 25, 33, 38–101](#)). Characteristics and critical appraisals of included records



are reported in **Supplementary Table S2**, and all following results are reported in terms of the number of records which used a given method. Most records ($n=63$) used the Actigraph with ambulatory PwMS during daily life, though two records monitored PA of wheelchair users and 12 used the devices during scripted, in-clinic walking tasks.

PA metrics included step counts ($n=26$), activity counts ($n=17$), sedentary behavior ($n=22$), time spent in LPA ($n=22$), time in MVPA ($n=36$), total time in PA ($n=1$), and energy expenditure ($n=1$), though two records did not report which metrics were used. Most records reported daily PA metrics ($n=63$), though they occasionally studied intra-day

patterns ($n=7$). The preprocessing filters, data sources, and cutpoints used to derive these PA metrics varied (Table 2). Records differed in their device placement (Hip: 43, Wrist: 7, Unspecified: 19), epoch lengths (1 s to 1 min, 26 unspecified) and sampling frequency settings (30 Hz–100 Hz, 44 unspecified). Most records' methods were not sufficiently reported to be reproducible ($n=64$). Forty-nine explicitly followed existing methods, but 42 of these had reporting gaps. Nine studies modified established methods without further validation. Common modifications included changing the filter used to derive cutpoint-based metrics ($n=5$) and using methods developed for hip sensor placement on wrist-derived data ($n=3$). Three records established new methods for deriving PA metrics. A searchable and filterable table detailing methods used by each study is provided in **Supplementary Table S2**.

Studies also varied in their wear time definitions. Most of the 63 records which described real-world PA (as opposed to in-clinic tasks) did not fully report their wear time definitions ($n=35$). Of those that did, studies defined inactive periods of 30 min ($n=5$), 60 min ($n=17$), 90 min ($n=3$), or 180 min ($n=3$) as non-wear time. Inactive periods were either defined as periods of continuous zeros ($n=18$), or calculated through two wear time algorithms: the Troiano algorithm ($n=8$) or the Choi algorithm ($n=2$). The Troiano algorithm allows for 2-minute spikes before an inactive period is broken, while the Choi algorithm allows for 2 min spikes but also requires 30 min of continuous zeros before and after spikes to qualify as a non-wear period (19, 107). One additional study defined valid days as those with 300 or more steps, and did not differentiate between wear and non-wear time. Minimum daily wear time ranged from 10 to 16 h (20 unspecified) and several studies ($n=16$) considered days invalid if they contained any non-wear periods. Studies also differed in the minimum number of valid days required to consider a participant's measurement valid (1 to 6 days, 31 unspecified), and whether they differentiated between weekdays and weekend days in their assessments of measurement validity (57 unspecified). This information is provided in greater detail in **Supplementary Table S2**.

3.1.2 Method selection

Based on review findings, we selected established methods and their common derivatives to quantify the potential effect of methodological heterogeneity on six PA metrics: step count, activity counts, total time in PA, sedentary time, time in light PA, and time in MVPA (24, 25, 33, 64, 85, 104–106). Due to the characteristics of our study population, we limited this comparison to methods which were designed for ambulatory individuals, rather than those who regularly use wheelchairs. In addition, we limited our analysis to methods designed for a hip sensor location as this was both the most

commonly-used location and the location for which most methods were designed.

In particular, we assessed the effects of preprocessing filter (Standard filter or LFE), data source (Vertical axis counts, VM counts, or in some cases step count), three cutpoints to differentiate between sedentary and non-sedentary behavior (100, 150, and 200 cpm), and seven methods to differentiate between MVPA and non-MVPA (Tables 3, 4). We also assessed the effect of wear-time definitions on these metrics, considering three wear time algorithms [continuous zeros (Zeros), Troiano, Choi] (19, 107) and four inactivity periods (30, 60, 90, and 180 min) which were commonly used by included studies. Other methodological factors, such as the definition and number of valid days required for a valid reading (Table 1), have been addressed previously and were not studied here (31, 32).

3.2 Experimental study

3.2.1 Participant characteristics

Of the 47 enrolled participants, 2 left rehabilitation early and were excluded from the study. One additional Actigraph data file was corrupted and two participants did not meet minimum wear time requirements, leaving 42 datasets for analysis. Participant characteristics are presented in Table 5. According to our primary non-wear time definition (60 min of continuous zeros), participants wore the Actigraph devices for a median of 13.7 h on 5.8 days. However, wear time estimates varied according to wear time algorithm and inactivity period length ($p < 0.001$), and estimates ranged from an average of 4.4 valid days with 12.1 h of wear time per day (Troiano algorithm, 30 min of inactivity) to 5.9 days with 14.2 h of wear time per day (180 min of continuous zeros).

3.2.2 Wear time

According to our primary non-wear time definition (60 min of continuous zeros), participants wore the Actigraph devices for a median of 13.7 h on 5.8 days. However, linear regression showed that wear time estimates were affected by choice of filter (Standard vs. LFE), wear time algorithm (Zeros, Choi, Troiano), and inactivity period (30, 60, 90, or 180 min). Filter and algorithm terms exhibited interactions with inactivity period. Specifically, the wear time estimates substantially decreased when a 30 min inactivity period was used with the Troiano algorithm, though use of the LFE decreased this effect. Model coefficients are provided in **Supplementary Table S3** for wear time and all PA metrics. Estimates ranged from an average of 4.4 valid days with 12.1 h of wear time per day (Troiano algorithm, 30 min of inactivity) to 5.9 days with 14.2 h of wear time per day (180 min of continuous zeros).

TABLE 2 Frequency table of the methods used to derive physical activity metrics identified in the rapid review.

	Step count (n = 26)	Activity count (n = 17)	Time in PA (n = 1)	Sedentary behavior (n = 22)	Time in LPA (n = 22)	Time in MVPA (n = 36)
Sensor location						
Hip	14	8	1	15	14	23
Wrist	3	5	0	2	2	6
Unspecified/Unclear	9	4	0	5	6	7
Preprocessing filter						
Standard (88)	0	0	0	0	0	1
LFE (88)	6	3	0	9	9	10
Unspecified/Unclear	20	14	1	13	13	25
Data source						
Vertical axis	–	3	0	15	15	20
Vector magnitude	–	9	0	5	5	7
Step counts	–	–	–	–	–	1
Unspecified/Unclear	–	5	1	2	2	8
Cutpoint – Sedentary behavior vs. LPA						
100 Vertical cpm (102)	–	–	0	15	15	–
150 VM cpm (102)	–	–	0	4	4	–
200 VM cpm	–	–	0	1	1	–
Unspecified/Unclear	–	–	1	2	2	–
Cutpoint – Light PA vs. MVPA						
1,584 vertical cpm ^a (24)	–	–	–	–	5	8
1,722 vertical cpm ^a (24)	–	–	–	–	5	5
1,745 vertical cpm ^a (23)	–	–	–	–	0	2
EDSS 0-5.5: 1,980 vertical cpm ^a (23)	–	–	–	–	2	4
EDSS 6.0+: 1,185 vertical cpm ^a						
1,952 vertical cpm (103)	–	–	–	–	3	3
2,690 VM cpm (104)	–	–	–	–	5	5
EDSS 0-3.5: 99 step/min (105)	–	–	–	–	–	1
EDSS 4.0-5.5: 89 steps/min						
EDSS 6.0+: 79 steps/min						
3,644 VM cpm (69) ^b	–	–	–	–	–	2
Unspecified/Unclear	–	–	–	–	2	8
Wear time algorithm						
Continuous zeros	8	4	0	6	6	7
Troiano	2	1	0	4	4	4
Choi	1	1	0	1	0	1
Not used/Not applicable	1	0	0	0	0	3
Unspecified/Unclear	11	11	1	11	12	21
Wear time inactivity period						
30 min	0	0	0	3	2	2
60 min	10	3	0	4	5	5
90 min	1	1	0	1	0	2
180 min	0	2	0	3	3	3
Not used/Not applicable	1	0	0	0	0	3
Unspecified/Unclear	11	11	1	11	12	21

The table reflects the number of records which reported using each method. Some records used multiple methods, therefore sums may exceed the total number of records using each PA metric.

^aMultiple cutpoints were developed in a single study. Often, studies reference the original paper without designating which cutpoint or Actigraph settings were used.

^bDerived for MVPA during wheelchair propulsion. All other cutpoints were derived for ambulatory individuals.

LFE, low frequency extension; VM, vector magnitude; cpm, counts per minute; PA, Physical activity; EDSS, Expanded disability status scale; LPA, Light physical activity; MVPA, Moderate to vigorous physical activity.

TABLE 3 Methodological factors with the potential to affect six PA metrics derived from the Actigraph GT3X.

	Step count	Activity counts	Time in PA	Sedentary time	Time in LPA	Time in MVPA
Filter	X	X	X	X	X ^a	X ^a
• Standard						
• LFE						
Data source	X	X	X	X	X ^a	X ^a
• Vertical axis						
• Vector magnitude						
• Steps						
Cutpoint – Sedentary			X	X	X	
• 100 cpm						
• 150 cpm						
• 200 cpm						
Cutpoint – MVPA					X	X
• Sandroff-1584						
• Sandroff-1722						
• Sandroff-1745						
• Sandroff-Severity						
• Freedson						
• Sasaki						
• Agiovlasis (Standard filter) ^b						
• Agiovlasis (LFE) ^b						
Wear time algorithm	X	X	X	X	X	X
• Zeros						
• Choi						
• Troiano						
Inactivity period	X	X	X	X	X	X
• 30 min						
• 60 min						
• 90 min						
• 180 min						

^aMVPA cutpoints are designed to be used with specific filters and data sources, therefore the independent effect of these factors on PA metrics which utilize MVPA cutpoints were not directly assessed. Rather, these factors are implicit to the employed MVPA cutpoints as described in Table 4.

^bStep count based methods are not available to differentiate between sedentary and non-sedentary minutes, therefore the Agiovlasis cutpoints cannot be used to estimate LPA.

LFE, low frequency extension; cpm, counts per minute; PA, Physical activity; LPA, light physical activity; MVPA, moderate to vigorous physical activity.

3.2.3 Step count

Linear mixed effects modeling showed that choice of filter (Standard vs. LFE) affected raw daily step count estimates, whereas wear time algorithm and inactivity period did not. MS severity acted as an effect modifier with respect to the impact of the filter; the LFE disproportionately increased step counts in severe MS compared to mild and moderate MS (Supplementary Table S3). When step counts were adjusted for wear time, they became independently affected by wear time algorithm and inactivity period, with the Troiano algorithm and a 30 min inactivity period imparting the largest biases.

Step counts derived from the two filtering methods differed ($p < 0.001$), with population medians of 5,852 and 11,695 steps per day when the Standard filter and LFE were used, respectively (Figure 2A). The correlation (Pearson's r) and agreement (CCC) between methods were 0.89 and 0.31, respectively (Supplementary Table S4).

3.2.4 Activity counts

Regression showed that preprocessing filter (Standard vs. LFE) and data source (Vertical axis vs. VM) independently

affected activity counts, but wear time algorithm and inactivity period did not. Activity counts decreased with increasing MS severity. Fully adjusted models identified interactions between filter and data source and effect modification due to disease severity, but these terms were not retained in reduced models. When adjusted for wear time, activity counts became independently affected by wear time algorithm and inactivity period. The Troiano algorithm and a 30 min inactivity period imparted the largest biases (Supplementary Table S3).

Unadjusted activity counts derived from the resulting four methods differed ($p < 0.001$) with population medians ranging from 179,324 to 533,753 counts per day (Figure 2B). In pairwise analysis ($n = 6$ pairs), all methods differed from one another (each $p < 0.001$) (Supplementary Table S4). Correlations between pairs ranged from 0.90 to 1.00. Agreement was high when methods used the same data source (CCC: 0.94–0.97) and very weak when data source differed (0.24–0.37). All pairwise differences, correlations, and agreement, including confidence intervals, are reported in Supplementary Table S4.

TABLE 4 Methods for deriving moderate to vigorous activity metrics in ambulatory people with multiple sclerosis.

Metric & Method	Filter	Data source	Cutpoint	Original target population
Sandhoff-1,584 (24)	Standard	Vertical axis	1584 cpm	Individuals with MS
Sandhoff-1,722 (24)	LFE	Vertical axis	1722 cpm	Individuals with MS
Sandhoff-1,745 (23)	LFE	Vertical axis	1745 cpm	Individuals with MS
Sandhoff-Severity (23)	LFE	Vertical axis	EDSS 0-5.5: 1980 cpm EDSS 6.0+: 1185 cpm	Individuals with MS
Freedson (103)	LFE	Vertical axis	1952 cpm	Healthy adults
Sasaki (102, 104)	Standard	VM	2690 cpm	Healthy adults
Agiovlasitis (105)	Standard or LFE	Steps	EDSS 0-3.5: 99 steps/min EDSS 4.0-5.5: 89 steps/min EDSS 6.0+: 79 steps/min	Individuals with MS

LFE, low frequency extension; VM, vector magnitude; cpm, counts per minute; PA, Physical activity; EDSS, Expanded disability status scale.

TABLE 5 Participant characteristics.

Participants	42
Age	46 [40–51]
Sex (Females)	27 (64.3)
EDSS score	4.5 [3.5–6]
Mild	13 (31)
Moderate	17 (40)
Severe	12 (29)
MSWS-12 score	52 [27–67]
6MWT (m) (<i>n</i> = 33)	345 [242–428]
No walking aid	22 (67)
2 Sticks	9 (27)
Rollator	1 (3)
Other	1 (3)
10 mWT (s) (<i>n</i> = 40)	9 [7–13]
No walking aid	26 (65)
2 Sticks	8 (20)
Rollator	2 (5)
Other	4 (10)
10 m Gait speed (<i>n</i> = 41)	
>1.0 m/s	25 (61)
0.6–0.9 m/s	11 (27)
<0.6 m/s	5 (12)

Characteristics reported as *N* (%) or Median [Q2–Q4].

3.2.5 Total time in physical activity

Preprocessing filter, data source, and sedentary cutpoint independently affected estimates of time in PA. The Troiano wear time algorithm and 30 min inactivity period yielded statistically significant, but negligible increases in time in PA. Regardless of method, time in PA decreased with increasing MS severity. The LFE increased estimates of time in PA compared to the standard filter, while increasing the sedentary cutpoint reduced time in PA. Though the use of VM counts increased time in PA for all levels of MS severity, it disproportionately increased time in PA for participants with

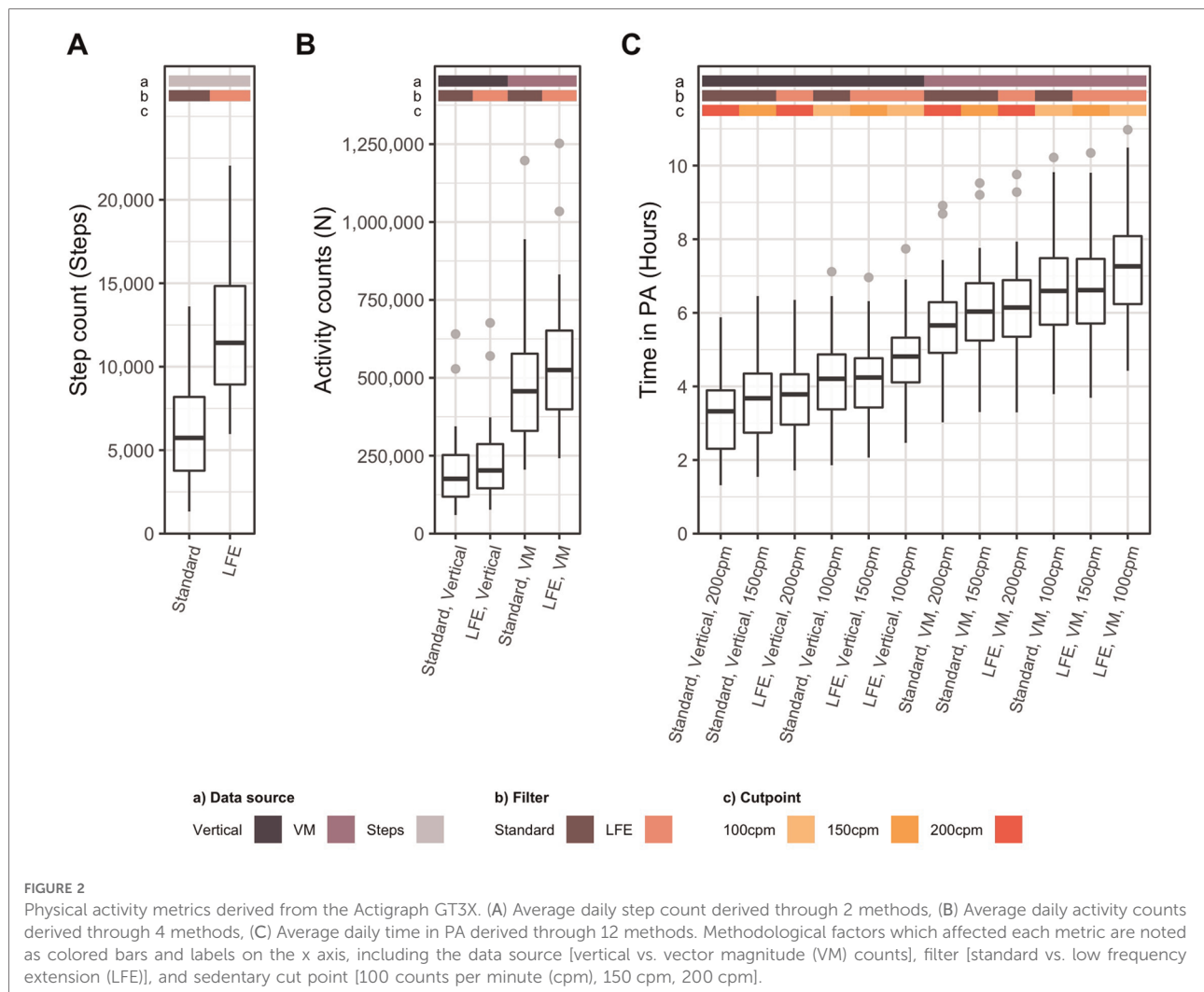
severe MS. When time in PA was adjusted for wear time, it became dependent on wear time algorithm and inactivity period. In this analysis, the inactivity period term also exhibited interactions with filter, data source, and wear time algorithm terms. These interactions were mostly attributed to the sensitivity 30 min inactivity period to other methods (**Supplementary Table S3**).

Unadjusted total time in PA derived from the 12 resulting methods differed ($p < 0.001$), with population medians ranging from 3.3 to 7.5 h of PA per day (**Figure 2C**). In pairwise analysis ($n = 66$ pairs), all pairs except two differed significantly from one another (each $p < 0.001$) (**Supplementary Table S4**). In these two cases, differences caused by the filter and sedentary cutpoint offset each other. Correlations between methods ranged from 0.61 to 1.00, while agreement ranged from 0.10 to 1.00. Agreement tended to be highest for pairs which (1) used the same data sources and (2) used a higher cutpoint with the LFE compared to the standard filter (**Supplementary Table S4**).

3.2.6 Sedentary behavior

Unlike other PA metrics, regression showed that the preprocessing filter, data source, cutpoint, wear time algorithm, and inactivity period definition all independently affected sedentary time estimates. Data source, wear time algorithm, and inactivity period exhibited interactions (i.e., effect modification) by disease severity. The effects of the filters exhibited interactions with wear time algorithm and inactivity period, and inactivity period and wear time algorithm exhibited interactions with each other. Adjusting for wear time reduced, but did not eliminate, the dependency of sedentary time on wear time algorithm and inactivity period (**Supplementary Table S3**).

Unadjusted sedentary time estimates stemming from the 120 possible combinations of methods significantly differed from each other ($p < 0.001$), ranging from 5.2 to 10.9 h of sedentary time per day (**Figure 3**). In pairwise analysis, 6,718



of the 7,140 pairs significantly differed from each other (**Supplementary Table S4**). Correlations between methods ranged from 0.15 to 1.00, whereas agreement (CCC) ranged from 0.03 to 0.99 (**Supplementary Table S4**).

3.2.7 Time in light physical activity

Time spent in LPA was independently affected by sedentary cutpoint and MVPA cutpoint. The Troiano wear time algorithm and 30 min inactivity period yielded statistically significant, but negligible increases in time in LPA. Estimates of LPA decreased with increasing sedentary cutpoint. Vertical axis-based MVPA cutpoints which used LFE (Sandroff 1,722, Sandroff 1,745, Sandroff Severity, Freedson) registered more time in LPA than the Sandroff-1,584 cutpoint, which uses the standard filter. The Sasaki cutpoint, which is based on VM counts, yielded the highest time LPA. Time in LPA decreased in moderate and severe MS compared to mild MS, though MS severity was an effect modifier for some MVPA cutpoints (Sandroff 1,584, Sasaki, Sandroff Severity). In addition to these dependencies,

adjustment for wear time made time in LPA dependent on wear time algorithm and inactivity period. These terms exhibited interactions both with each other, and multiple other methodological factors (**Supplementary Table S3**).

The 18 methods of calculating time in LPA significantly differed from each other ($p < 0.001$), with population medians ranging from 2.8 to 6.1 h per day (**Figure 4A**). In pairwise analysis, 141 of the 153 resulting pairs significantly differed from each other ($p < 0.001$) (**Supplementary Table S4**). Correlations between methods ranged from 0.61 to 1.00 and agreement ranged from 0.09 to 1.00. The Sasaki method, which was the only VM-based method, consistently exhibited the lowest correlations and agreement with all other vertical axis based methods.

3.2.8 Time in moderate to vigorous physical activity

Cutpoint, but not wear time algorithm or inactivity period definition, independently affected estimates of time in MVPA.

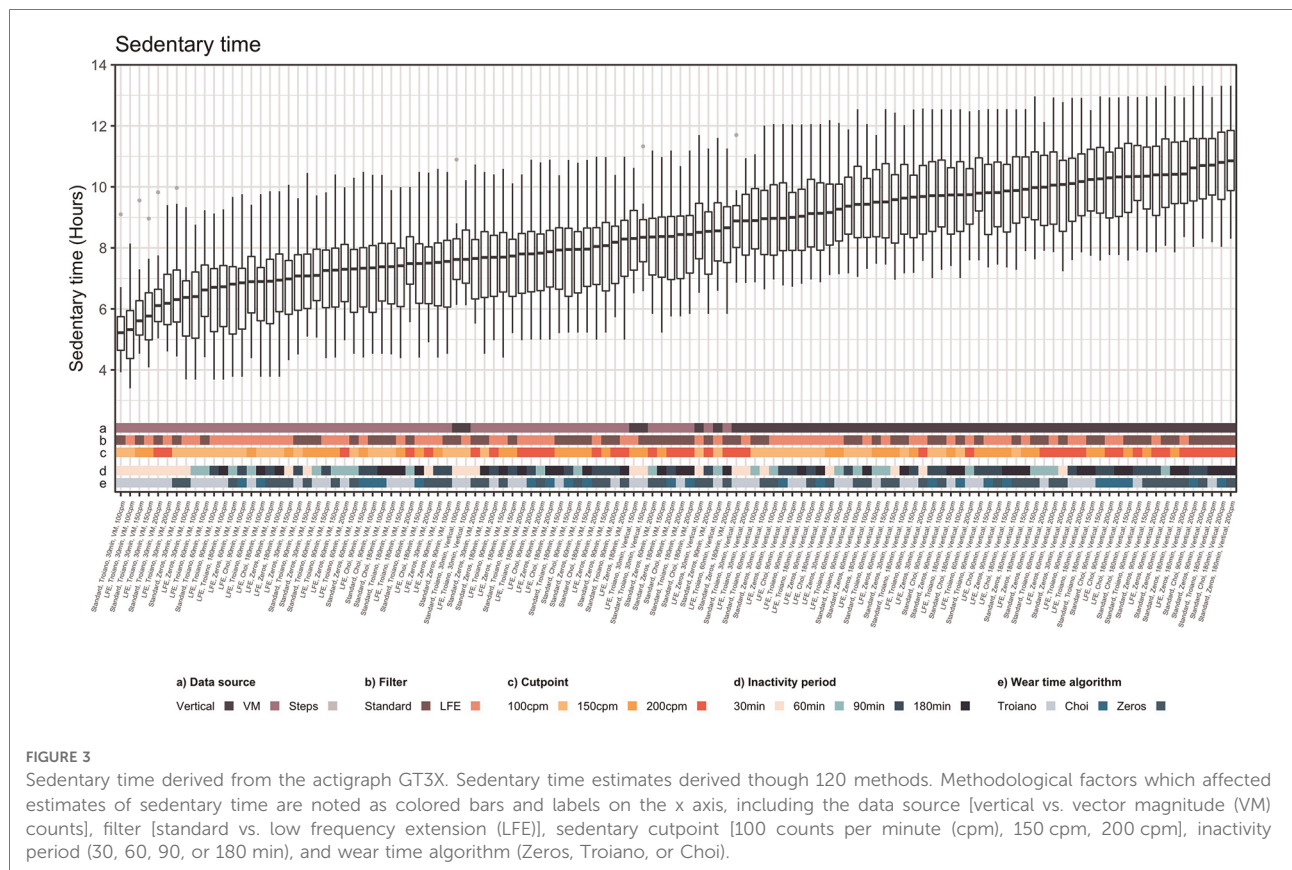


FIGURE 3

Sedentary time derived from the actigraph GT3X. Sedentary time estimates derived through 120 methods. Methodological factors which affected estimates of sedentary time are noted as colored bars and labels on the x axis, including the data source [vertical vs. vector magnitude (VM) counts], filter [standard vs. low frequency extension (LFE)], sedentary cutpoint [100 counts per minute (cpm), 150 cpm, 200 cpm], inactivity period (30, 60, 90, or 180 min), and wear time algorithm (Zeros, Troiano, or Choi).

Vertical axis based methods which were developed for MS populations (Sandroff 1,584, Sandroff 1,722, Sandroff 1,745, Sandroff Severity) did not differ from each other at the population level. The Freedson method, which was developed for healthy adults, and the Agiovaslitis method, which was based on step cadence, registered less time in MVPA than these methods. The Sasaki method, which was based on VM counts, yielded the highest time in MVPA. Time in MVPA decreased in moderate and severe MS compared to mild MS, but MS severity was not an effect modifier. Adjustment for wear time caused statistically significant, though negligible increases to time in MVPA (Supplementary Table S3).

The eight methods of calculating time in MVPA significantly differed from each other ($p < 0.001$), ranging from population medians of 5 to 36 min per day (Figure 4B). Twenty-two of the 27 pairs significantly differed from each other in pairwise analysis (Supplementary Table S4). The four vertical axis based methods (Sandroff-1584, Sandroff-1722, Sandroff-1745, Freedson) demonstrated almost perfect correlation and agreement with each other ($r: 0.99-1.00$, CCC: 0.96-1.00), and pairings of these methods and the Sasaki method (VM-based) or the Sandroff-Severity method (applied different cutpoints across MS severity strata) also exhibited high correlation and agreement ($r: 0.79-0.91$; CCC: 0.72-0.88). The Agiovaslitis method, which was based on step

cadence, consistently exhibited the lowest correlation and agreement with other methods ($r: 0.59-0.79$, CCC: 0.23-0.52) (Supplementary Table S4).

3.2.9 How can methodological heterogeneity bias PA research?

3.2.9.1 Meeting PA recommendations

Using this study as an illustrative example, 92.9% of participants would have met a 7,500 step target using the LFE, whereas only 33.3% would have met this threshold with the standard filter (Table 6). Similarly, the percentage of the population which met an MVPA target of 30 min per day ranged from 14.3% with the Agiovaslitis method with the standard filter to 59.5% using the Sasaki method (Table 6).

3.2.9.2 Relationships between PA and clinical outcomes

Correlations with clinical measures varied across methods for all PA metrics (Supplementary Table S5). Both step count methods exhibited moderate to strong correlations with the EDSS, 6 MWT, 10 mWT, and MSWS-12, though point estimates tended to be higher for the standard filter than for the LFE. For all other metrics, vertical axis based methods tended to exhibit stronger correlations with clinical measures than VM based methods. Time in LPA did not consistently correlate with clinical measures via any method. No

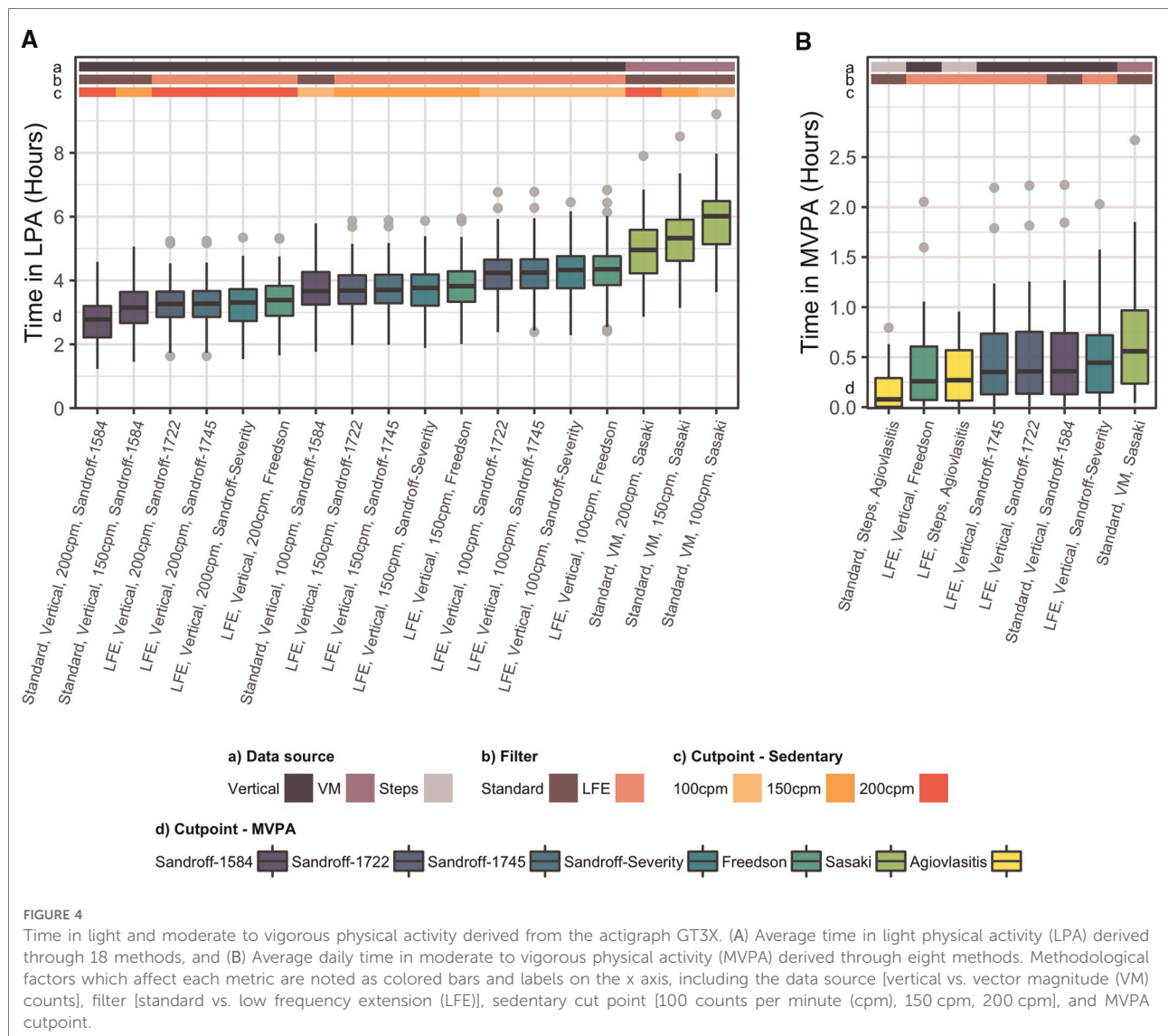


FIGURE 4

Time in light and moderate to vigorous physical activity derived from the actigraph GT3X. (A) Average time in light physical activity (LPA) derived through 18 methods, and (B) Average daily time in moderate to vigorous physical activity (MVPA) derived through eight methods. Methodological factors which affect each metric are noted as colored bars and labels on the x axis, including the data source [vertical vs. vector magnitude (VM) counts], filter [standard vs. low frequency extension (LFE)], sedentary cut point [100 counts per minute (cpm), 150 cpm, 200 cpm], and MVPA cutpoint.

TABLE 6 Percentage of study participants meeting physical activity guidelines according to multiple methods.

Method	% meeting PA guidelines
7,500 steps per day	
Standard filter	33.3%
LFE	92.9%
30 min of MVPA per day	
Sandroff 1584	40.5%
Sandroff 1722	40.5%
Sandroff 1745	38.1%
Freedson	31.0%
Sasaki	59.5%
Agiovlasisitis (Standard filter)	14.3%
Agiovlasisitis (LFE)	26.2%

objectively-measured PA metric correlated with the IPAQ. Adjustment for wear time did not consistently affect the associations between PA metrics and clinical outcome measures.

4 Discussion

In this rapid review and comparative analysis, we found that existing literature on PA in MS populations exhibited high methodological heterogeneity, and methods were often poorly reported. Through direct comparison, we found that all PA metrics were sensitive to methodological heterogeneity, though the degree and uniformity of this sensitivity varied. Heterogeneity affected the outcome of analyses relevant to the MS literature, demonstrating potential confounding effects on

population-level findings and meta-analyses. Though this study was not designed to assess the relative accuracy or validity of included methods and metrics, our results suggest several PA metrics and methods which should be used and interpreted with caution in MS populations. We therefore make the following recommendations to guide future research with the Actigraph GT3X:

- Improved reporting and adherence to validated methods are needed
- Vertical axis counts may be preferable to VM counts for all PA metrics
- Step counts derived from hip-worn Actigraph GT3X are not recommended
- Time in PA may be preferable to sedentary time as an indicator of sedentary behavior
- Sedentary cutpoints require further research to accurately characterize light PA
- MVPA derived from different data sources are not comparable
- The Troiano algorithm and 30 min inactivity periods are not recommended

4.1 Improved reporting and adherence to validated methods are needed

Our comparative analysis demonstrated that methodological factors have the potential to confound common PA metrics. However, we also showed that the reporting of Actigraph data processing methods is both limited and inconsistent in existing literature. These findings demonstrate that our ability to interpret and compare studies is limited, and highlight the need for improved methodological reporting in studies which use the Actigraph GT3X. At a minimum, authors should report the data source, filter, wear time algorithm, cutpoints, and sensor placement they used to derive their PA metrics. We also identified several instances in which established methods were altered when deriving PA metrics (eg, using the LFE when the method was developed for use with the standard filter). We have shown that these alterations may affect the validity of the resulting PA metrics, and recommend that authors adhere to the filters, sensor placements, and data sources for which specific cutpoints were developed. Finally, given the relatively large impact of methodological factors on study outcomes, data processing methods should also be specified a-priori.

4.2 Vertical axis counts may be preferable to VM counts for all PA metrics

Of the methodological factors assessed here, discrepant data sources consistently yielded the largest differences in PA estimates. PA metrics derived from vertical axis and VM counts

exhibited low agreement and correlation with each other, with VM counts yielding significantly higher PA estimates than vertical axis based methods. Metrics derived from these data sources behaved differently across disease severity strata, with VM based metrics disproportionately increasing PA estimates in those with severe MS. As a result, vertical axis based methods consistently exhibited stronger correlations with clinical measures, suggesting superior construct validity. We therefore do not recommend direct comparisons between studies which derive PA from different data sources, and recommend the use of vertical axis based methods over VM based methods for all PA metrics evaluated here.

4.3 Step counts derived from hip-worn Actigraph GT3X are not recommended

Daily step count was one of the most common PA metrics studied in our review, though methods used to derive step count were not commonly reported. In our review, only six of the 26 studies which measured step count reported whether they used the standard filter or the LFE. However, we showed that filter selection strongly affected step count estimates, with the LFE nearly doubling step counts relative to the standard filter. This effect was inconsistent across disease severity strata, disproportionately increasing PA estimates in those with severe MS.

The manufacturer of the Actigraph recommends using the LFE in populations which move slowly or take very light steps, though they do not define a specific threshold below which the LFE should be used (16). In an independent study, Bezuidenhout et al. showed that the standard filter and LFE perform similarly when walking speeds exceed 1.0 m/s during scripted, over-ground walking tasks, whereas the LFE is more accurate at walking speeds below 1.0 m/s (18). However, Feito et al. showed that the standard filter underestimates step count by 25%–30% and the LFE overestimates step count by 30% when daily step counts are evaluated under free-living conditions (108). Feito et al. hypothesized that these biases were due to reduced sensitivity to slow walking and mischaracterization of non-walking movement, respectively (108).

In this study, 16 of the 42 participants (38%) exhibited normal walking speeds slower than 1.0 m/s, suggesting that the LFE may be preferable in our population based on Bezuidenhout et al.'s findings (18). However, like Feito et al., we found that step counts derived with the LFE were unrealistically high for those of all MS severity levels (108). In addition, we found that step counts derived using the standard filter exhibited stronger correlations with clinical measures. However, due to the standard filters' propensity to underestimate step count in slow walkers, it is probable that these relationships are confounded by the relationship between MS severity and walking speed. For these reasons, we caution researchers against using an Actigraph GT3X worn on the hip

to measure step counts. If step counts are desired as PA outcomes, other devices or ankle placement of the Actigraph have been shown to provide higher accuracy (18, 108, 109).

4.4 Time in PA may be preferable to sedentary time as an indicator of sedentary behavior

Sedentary behavior is a relevant outcome in MS populations (110) and was the second most frequently studied PA metric in our review. However, we found that sedentary time estimates are highly sensitive to data processing methodology and wear time definitions, consistent with findings in a healthy population (111). Adjustment for daily wear time has been used to account for the relationship between sedentary time and wear time, reducing inter-personal variability arising from differences in wear time (21, 38–40). We found that such adjustment reduced, but did not eliminate, the biases imparted by sedentary time's dependency on wear time or other methodological factors. In addition, the methods affecting sedentary time estimates exhibited interactions and were poorly reported in the literature. This reporting gap and the multiple interactions exhibited between methodological factors in our models make it nearly impossible to account for independent effects of individual factors or control for differences across studies.

We found that total time in PA, which is effectively the inverse of sedentary time, is not sensitive to wear time definition and exhibits fewer interactions between methodological factors. Though our rapid review showed that time in PA is not widely used as an outcome in the MS literature, it may nevertheless be a preferable, albeit indirect, indicator of sedentary behavior until the relative validity of sedentary time methods is clarified. In addition, we recommend using vertical axis counts rather than VM counts to derive time in PA, as this method consistently exhibited stronger relationships with clinical measures.

4.5 Sedentary cutpoints require further research to accurately characterize light PA

Kozy Keadle et al. previously assessed the validity of multiple sedentary cutpoints in overweight adults using vertical axis counts and the LFE, recommending 150 cpm in that population with those sensor settings (112). In a healthy population, Carr and Mahar found that 100 cpm is an appropriate cutpoint for single-axis measurements with the Actigraph GT1M (113), which is roughly equivalent to the vertical axis based measures using the LFE on the GT3X (25). They also found that a cutpoint of 150cpm provides higher

accuracy in VM based measurements without the LFE (113). However, both of these studies aimed to assess accuracy with respect to sedentary behavior, and both cutpoints exhibited lower accuracy when identifying light physical activity. We cannot recommend one cutpoint over the others based on this study. However, given our recommendation to use time in PA rather than sedentary time, we do recommend further study of sedentary cutpoints to identify thresholds which accurately identify time in light physical activity. During this research, cutpoints should be confirmed for both the standard filter and the LFE. We also suggest that researchers more readily report their cutpoints, data sources, and filters to enable comparisons across studies.

4.6 MVPA derived from vertical axis, VM, and step counts are not comparable

Time in MVPA was the most frequently-used PA metric identified by our rapid review. Though eight methods were used to derive MVPA, four methods derived from vertical magnitude counts performed similarly (Sandroff-1584, Sandroff-1722, Sandroff-1754, and Freedson). Three of these methods – Sandroff-1584, Sandroff-1722, and Sandroff-1754 – were essentially equivalent and exhibited strong, consistent correlations with clinical measures. The Freedson method employed a higher cutpoint originally developed for a healthy population, and therefore yielded lower estimations of PA (104). However, it yielded nearly identical correlations with clinical measures as the other vertical axis based methods. Together, these four methods were used in the majority of records identified in our review.

The Sandroff-Severity method applies different cutpoints to data from individuals with mild, moderate, and severe MS, as individuals with severe MS exert more energy than those with mild and moderate MS to complete the same tasks, thereby lowering the threshold at which activities qualify as “moderate to vigorous” PA (24, 33, 106). It is therefore unsurprising that this method demonstrated lower agreement and lower correlations with clinical measures compared to other methods. Sandroff et al. derived these cutpoints for sub-populations in MS, but did not necessarily intend for them to be used as a single method (24). It is currently unclear which of these approaches – uniform cutpoints or severity-specific cutpoints – yield more valid population-level and individual-level representations of MVPA. Direct comparisons with a gold standard may be beneficial in future research.

The Agiovlasis method, which is based on step cadence, yielded the lowest estimations of MVPA and exhibited the lowest correlation and agreement with other methods. MVPA estimates made though this method were highly sensitive to choice of filter, since they were derived from step counts. However, both implementations of this method yielded similar correlations with clinical measures as the vertical axis

based measures. This method is therefore not directly comparable to vertical axis based methods, but nevertheless demonstrates evidence of construct validity.

The Sasaki method, which is based on VM counts, exhibited only moderate correlations with clinical measures, moderate correlations with other methods, and yielded by far the highest estimate of MVPA time. Similar to other PA metrics derived from VM counts, we cannot recommend the use of the Sasaki method until additional research establishes its validity.

4.7 The Troiano algorithm and 30 min inactivity periods are not recommended

When PA metrics were affected by wear time methods, they were consistently the most strongly biased by the Troiano algorithm and the use of a 30 min inactivity period. These methodological options enable shorter inactivity bouts, with or without activity spikes, to be counted as non-wear time. While this bias only affected sedentary time estimates in unadjusted PA metrics, it affected all PA metrics which were adjusted for wear time. The Troiano algorithm continues to be widely used in the physical activity literature (15), though it has been found to substantially underestimate wear time when it is used with hip-worn Actigraphs (107). This is consistent with our own findings, as the wear time estimates produced by the Troiano algorithm were substantially lower than those produced by the Choi algorithm or continuous zeros. The remaining two algorithms – Choi algorithm and continuous zeros – were not equivalent. However, they imparted smaller relative biases on both unadjusted and adjusted PA metrics compared to the Troiano algorithm. For these reasons, we recommend that researchers adopt either the Choi algorithm or the continuous zeros method to calculate wear time. We further recommend that wear time is calculated with an inactivity period of at least 60 min due to the level of bias induced and the number of interactions exhibited by the 30 min wear period.

4.8 Future research

These findings highlight that further research is needed to establish both the absolute and relative validity, reliability, responsiveness of PA metrics derived through different data processing methods. This research should include not only clinical assessments of scripted PA, but also validation compared to a reference standard during free-living assessments (114), as many of the methods studied here were developed and validated only during scripted walking tasks in clinical settings (23, 32, 69, 104). Once the relative validity, reliability, and responsiveness of these methods are established, an optimal set of PA metrics derived from the Actigraph GT3X can be defined for MS populations.

4.9 Strengths and limitations

This study reviews existing literature, quantifies methodological heterogeneity associated with the most widely-used PA monitor in MS, and illustrates the potential consequences of this heterogeneity on PA metrics through direct, pairwise comparisons. We offer a comprehensive exploration of the impact of methodological heterogeneity on Actigraph-based PA metrics in MS populations.

However, this study is not without its limitations. The summary of the existing literature is based on a rapid review rather than a systematic review, and therefore made compromises to ensure feasibility. However, the methods identified in the review appeared to reach saturation, defined as the point at which new data are unlikely to alter conclusions, and it is unlikely that we missed any commonly-used data processing methods. For PA metrics which rely on cutpoints, such as LPA and MVPA, we found that studies often deviated from the filters and data sources for which specific cutpoints were developed. We did not assess the effects of every possible combination of filter, data source, and cutpoint on these metrics. However, deviations from established methods are likely to further increase variability in PA literature. In our analysis, PA metrics were derived through various methods from the same raw data, introducing dependencies which could introduce overfitting and bias standard errors or *p* values. However, we mitigated this risk by adjusting for participant-level random effects, using BIC values (rather than *p* values or AIC) to guide model development, using paired analysis methods where appropriate, and adjusting analyses for multiple testing. Quantitative comparisons were conducted based on data collected during and immediately after inpatient rehabilitation, which included structured exercise training and physiotherapy. The PA behavior described here is therefore not necessarily representative of free-living PA. Finally, PA monitors other than the Actigraph GT3X are not considered here, and care should be taken when comparing findings between devices.

5 Conclusions

Though the Actigraph GT3X is widely used and is often considered well-validated in MS populations, the PA metrics derived from this device are sensitive to methodological choices during data processing. Methodological heterogeneity of existing literature is high, and additional research is needed to establish the relative validity and responsiveness of existing methods. Researchers should consider the impact of methodological heterogeneity on PA metrics when selecting and reporting their methods, as these decisions may limit the responsiveness, interpretability, and cross-study comparability of PA outcomes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics committee of the canton of Zurich (2020-02350). The patients/participants provided their written informed consent to participate in this study.

Author contributions

AP contributed to study design, data analysis, interpretation of the results, and drafting of the manuscript. CS and CH contributed to study design, data collection, study management, data analysis, and interpretation of the results. JK contributed to study design, data collection, and study management. RS contributed to data collection and study management. VvW and RG contributed to study design and supervised all contributions. All authors contributed to the article and approved the submitted version.

References

- Motl RW, Sandroff BM, Kwakkel G, Dalgas U, Feinstein A, Heesen C, et al. Exercise in patients with multiple sclerosis. *Lancet Neurol.* (2017) 16(10):848–56. doi: 10.1016/S1474-4422(17)30281-8
- Motl RW, Pilutti LA. The benefits of exercise training in multiple sclerosis. *Nat Rev Neurol.* (2012) 8:487–97. doi: 10.1038/nrneurol.2012.136
- Kalb R, Brown TR, Coote S, Costello K, Dalgas U, Garmon E, et al. Exercise and lifestyle physical activity recommendations for people with multiple sclerosis throughout the disease course. *Mult Scler J.* (2020) 26(12):1459–69. doi: 10.1177/1352458520915629
- Beckerman H, de Groot V, Scholten MA, Kempen JCE, Lankhorst GJ. Physical activity behavior of people with multiple sclerosis: understanding how they can become more physically active. *Phys Ther.* (2010) 90(7):1001–13. doi: 10.2522/ptj.20090345
- Khan F, Amatya B. Rehabilitation in multiple sclerosis: a systematic review of systematic reviews. *Arch Phys Med Rehabil.* (2017) 98:353–67. doi: 10.1016/j.apmr.2016.04.016
- Sandroff BM, Klaren RE, Pilutti LA, Dlugonski D, Benedict RHB, Motl RW. Randomized controlled trial of physical activity, cognition, and walking in multiple sclerosis. *J Neurol.* (2014) 261(2):363–72. doi: 10.1007/s00415-013-7204-8
- Block VAJ, Pitsch E, Tahir P, Cree BAC, Allen DD, Gelfand JM. Remote physical activity monitoring in neurological disease: a systematic review. *PLoS ONE.* (2016) 11:e0154335. doi: 10.1371/journal.pone.0154335
- Charron S, McKay KA, Tremlett H. Physical activity and disability outcomes in multiple sclerosis: a systematic review (2011–2016). *Mult Scler Relat Disord.* (2018) 20:169–77. doi: 10.1016/j.msard.2018.01.021
- Hur SA, Guler SA, Khalil N, Camp PG, Guenette JA, Swigris JJ, et al. Minimal important difference for physical activity and validity of the international physical activity questionnaire in interstitial lung disease. *Ann Am Thorac Soc.* (2019) 16(1):107–15. doi: 10.1513/AnnalsATS.201804-265OC
- Motl RW, Bollaert RE, Sandroff BM. Validation of the godin leisure-time exercise questionnaire classification coding system using accelerometry in multiple sclerosis. *Rehabil Psychol.* (2018) 63(1):77–82. doi: 10.1037/rep0000162
- Prince SA, Reid RD, Bernick J, Clarke AE, Reed JL. Single versus multi-item self-assessment of sedentary behaviour: a comparison with objectively measured sedentary time in nurses. *J Sci Med Sport.* (2018) 21(9):925–9. doi: 10.1016/j.jsams.2018.01.018
- Healy GN, Clark BK, Winkler EAH, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. *Am J Prev Med.* (2011) 41:216–27. doi: 10.1016/j.amepre.2011.05.005
- Harada ND, Chiu V, King AC, Stewart AL. An evaluation of three self-report physical activity instruments for older adults. *Med Sci Sports Exerc.* (2001) 33(6):962–70. doi: 10.1097/00005768-200106000-00016
- Heesch KC, van Uffelen JGZ, Hill RL, Brown WJ. What do IPAQ questions mean to older adults? Lessons from cognitive interviews. *Int J Behav Nutr Phys Act.* (2010) 7(1):35. doi: 10.1186/1479-5868-7-35
- Miguel JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Med.* (2017) 47:1821–45. doi: 10.1007/s40279-017-0716-0
- Actigraph Corp. GT3X+ and wGT3X+ Device Manual v2.0.0. 2013. p. 15 ps.
- Korpan SM, Schafer JL, Wilson KCS, Webber SC. Effect of actigraph GT3X+ position and algorithm choice on step count accuracy in older adults. *J Aging Phys Act.* (2015) 23(3):377–82. doi: 10.1123/japa.2014-0033
- Bezuidenhout L, Thurston C, Hagströmer M, Moulæe Conradsson D. Validity of hip and ankle worn actigraph accelerometers for measuring steps as a function of gait speed during steady state walking and continuous turning. *Sensors.* (2021) 21(9):3154. doi: 10.3390/s21093154

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fresc.2022.989658/full#supplementary-material>.

19. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. *Med Sci Sports Exerc.* (2012) 44(10):2009–16. doi: 10.1249/MSS.0b013e318258cb36
20. Sagawa Y, Watelain E, Moulin T, Decavel P. Physical activity during weekdays and weekends in persons with multiple sclerosis. *Sensors (Basel).* (2021) 21(11):3617. doi: 10.3390/s21113617
21. Aadland E, Ylvisäker E. Reliability of objectively measured sedentary time and physical activity in adults. *PLoS ONE.* (2015) 10(7):e0133296. doi: 10.1371/journal.pone.0133296
22. Tricco AC, Langlois EV, Straus SE. *Rapid reviews to strengthen health policy and systems: A practical guide.*, Geneva: World Health Organization, Vol. 385 (2017). 1–116.
23. Tricco AC, Antony J, Zarin W, Striffler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. *BMC Med.* (2015) 13(1):224. doi: 10.1186/s12916-015-0465-6
24. Sandroff BM, Motl RW, Suh Y. Accelerometer output and its association with energy expenditure in persons with multiple sclerosis. *J Rehabil Res Dev.* (2012) 49(3):467–76. doi: 10.1682/JRRD.2011.03.0063
25. Sandroff BM, Motl RW. Comparison of ActiGraph activity monitors in persons with multiple sclerosis and controls. *Disabil Rehabil.* (2013) 35(9):725–31. doi: 10.3109/09638288.2012.707745
26. von Wyl V, Gonzenbach R. Barriers to Physical Activity in People With MS (BarkA-MS). 2021 (cited 2022 Apr 1). Available at: <https://clinicaltrials.gov/ct2/show/study/NCT04746807>
27. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* (1983) 33(11):1444–52. doi: 10.1212/WNL.33.11.1444
28. Kempen J, De Groot V, Knol DL, Polman CH, Lankhorst GJ, Beckerman H. Community walking can be assessed using a 10-metre timed walk test. *Mult Scler J.* (2011) 17(8):980–90. doi: 10.1177/1352458511403641
29. Goldman MD, Marrie RA, Cohen JA. Evaluation of the six-minute walk in multiple sclerosis subjects and healthy controls. *Mult Scler.* (2008) 14(3):383–90. doi: 10.1177/1352458507082607
30. Goldman MD, Ward MD, Motl RW, Jones DE, Pula JH, Cadavid D. Identification and validation of clinically meaningful benchmarks in the 12-item multiple sclerosis walking scale. *Mult Scler.* (2017) 23(10):1405–14. doi: 10.1177/1352458516680749
31. Klaren RE, Hubbard EA, Zhu W, Motl RW. Reliability of accelerometer scores for measuring sedentary and physical activity behaviors in persons with multiple sclerosis. *Adapt Phys Act Q.* (2016) 33(2):195–204. doi: 10.1123/APAQ.2015-0007
32. Motl RW, Zhu W, Park Y, McAuley E, Scott JA, Snook EM. Reliability of scores from physical activity monitors in adults with multiple sclerosis. *Adapt Phys Act Q.* (2007) 24(3):245–53. doi: 10.1123/apaq.24.3.245
33. Sandroff BM, Riskin BJ, Agiovlasis S, Motl RW. Accelerometer cut-points derived during over-ground walking in persons with mild, moderate, and severe multiple sclerosis. *J Neurol Sci.* (2014) 340(1–2):50–7. doi: 10.1016/j.jns.2014.02.024
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc.* (1997) 57(1):289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
35. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* (1989) 45(1):255. doi: 10.2307/2532051
36. Klaren RE, Sebastiao E, Chiu CY, Kinnett-Hopkins D, McAuley E, Motl RW. Levels and rates of physical activity in older adults with multiple sclerosis. *Aging Dis.* (2016) 7(3):278–84. doi: 10.14336/AD.2015.1025
37. Klaren RE, Motl RW, Dlugonski D, Sandroff BM, Pilutti LA. Objectively quantified physical activity in persons with multiple sclerosis. *Arch Phys Med Rehabil.* (2013) 94(12):2342–8. doi: 10.1016/j.apmr.2013.07.011
38. Blikman LJ, Van Meeteren J, Horemans HL, Kortenhorst IC, Beckerman H, Stam HJ, et al. Is physical behavior affected in fatigued persons with multiple sclerosis? *Arch Phys Med Rehabil.* (2015) 96(1):24–9. doi: 10.1016/j.apmr.2014.08.023
39. Blikman LJM, Van Meeteren J, Rizopoulos D, De Groot V, Beckerman H, Stam HJ, et al. Physical behaviour is weakly associated with physical fatigue in persons with multiple sclerosis-related fatigue. *J Rehabil Med.* (2018) 50(9):821–7. doi: 10.2340/16501977-2375
40. Braakhuis HEM, Berger MAM, Van Der Stok GA, Van Meeteren J, De Groot V, Beckerman H, et al. Three distinct physical behavior types in fatigued patients with multiple sclerosis. *J Neuroeng Rehabil.* (2019) 16(1):1–9. doi: 10.1186/s12984-018-0454-z
41. Block VJ, Lizee A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *J Neurol.* (2017) 264(2):316–26. doi: 10.1007/s00415-016-8334-6
42. Block VJ, Zhao C, Hollenbach JA, Olgin JE, Marcus GM, Pletcher MJ, et al. Validation of a consumer-grade activity monitor for continuous daily activity monitoring in individuals with multiple sclerosis. *Mult Scler J - Exp Transl Clin.* (2019) 5(4):205521731988866. doi: 10.1177/2055217319888660
43. Bollaert RE, Motl RW. Physical and cognitive functions, physical activity, and sedentary behavior in older adults with multiple sclerosis. *J Geriatr Phys Ther.* (2019) 42(4):304–12. doi: 10.1519/JPT.0000000000000163
44. Bollaert RE, Motl RW. Self-efficacy and physical and cognitive function in older adults with multiple sclerosis. *Int J MS Care.* (2019) 21(2):63–9. doi: 10.7224/1537-2073.2018-001
45. Cederberg KL, Motl RW, McAuley E. Physical activity, sedentary behavior, and physical function in older adults with multiple sclerosis. *J Aging Phys Act.* (2018) 26(2):177–82. doi: 10.1123/japa.2016-0358
46. Cederberg KLJ, Jeng B, Sasaki JE, Bralley TJ, Walters AS, Motl RW. Physical activity, sedentary behavior, and restless legs syndrome in persons with multiple sclerosis. *J Neurol Sci.* (2019) 407:116531. doi: 10.1016/j.jns.2019.116531
47. Jeng B, Sasaki JE, Cederberg KL, Motl RW. Sociodemographic and clinical correlates of device-measured sedentary behaviour in multiple sclerosis. *Disabil Rehabil.* (2021) 43(1):42–8. doi: 10.1080/09638288.2019.1614683
48. Motl RW, Sasaki JE, Cederberg KL, Jeng B. Social-cognitive theory variables as correlates of sedentary behavior in multiple sclerosis: preliminary evidence. *Disabil Health J.* (2019) 12(4):622–7. doi: 10.1016/j.dhjo.2019.05.002
49. Motl RW, Sasaki JE, Cederberg KL, Jeng B. Validity of sitting time scores from the international physical activity questionnaire-short form in multiple sclerosis. *Rehabil Psychol.* (2019) 64(4):463. doi: 10.1037/rep0000280
50. Motl RW, Sandroff BM, Wingo BC, McCroskey J, Pilutti LA, Cutter GR, et al. Phase-III, randomized controlled trial of the behavioral intervention for increasing physical activity in multiple sclerosis: project BIPAMS. *Contemp Clin Trials.* (2018) 71(June):154–61. doi: 10.1016/j.cct.2018.06.017
51. Neal WN, Cederberg KL, Jeng B, Sasaki JE, Motl RW. Is symptomatic fatigue associated with physical activity and sedentary behaviors among persons with multiple sclerosis? *Neurorehabil Neural Repair.* (2020) 34(6):505–11. doi: 10.1177/1545968320916159
52. Cederberg KLJ, Jeng B, Sasaki JE, Sikes EM, Cutter G, Motl RW. Physical activity and self-reported sleep quality in adults with multiple sclerosis. *Disabil Health J.* (2021) 14(4):101133. doi: 10.1016/j.dhjo.2021.101133
53. Cleland BT, Papanek P, Ingraham BA, Harkins A, Garnier-Villareal M, Woo D, et al. Determinants of low bone mineral density in people with multiple sclerosis: role of physical activity. *Mult Scler Relat Disord.* (2020) 38(August 2019):101864. doi: 10.1016/j.msard.2019.101864
54. Dlugonski D, Pilutti LA, Sandroff BM, Suh Y, Balantrapu S, Motl RW. Steps per day among persons with multiple sclerosis: variation by demographic, clinical, and device characteristics. *Arch Phys Med Rehabil.* (2013) 94(8):1534–9. doi: 10.1016/j.apmr.2012.12.014
55. Ehling R, Bsteh G, Muehlbacher A, Hermann K, Brenneis C. Ecological validity of walking capacity tests following rehabilitation in people with multiple sclerosis. *PLoS ONE.* (2019) 14(8):1–12. doi: 10.1371/journal.pone.0220613
56. Eldemir K, Guclu-Gunduz A, Ozkul C, Eldemir S, Soke F, Irkec C. Associations between fatigue and physical behavior in patients with multiple sclerosis with no or minimal disability. *Fatigue Biomed Heal Behav.* (2021) 9(2):69–78. doi: 10.1080/21641846.2021.1923995
57. Engelhard MM, Patek SD, Lach JC, Goldman MD. Real-world walking in multiple sclerosis: separating capacity from behavior. *Gait Posture.* (2018) 59:211–6. doi: 10.1016/j.gaitpost.2017.10.015
58. Abonie US, Saxton J, Baker K, Hettinga FJ. Objectively-assessed physical activity and self-reported activity pacing in adults with multiple sclerosis: a pilot study. *Clin Rehabil.* (2021) 35(12):1781–8. doi: 10.1177/02692155211024135
59. Gulde P, Hermsdörfer J, Rieckmann P. Inpatient rehabilitation: prediction of changes in sensorimotor performance in multiple sclerosis: a pilot study. *J Clin Med.* (2021) 10(10):2177. doi: 10.3390/jcm10102177
60. Karle V, Hartung V, Ivanovska K, Mäurer M, Flachenecker P, Pfeifer K, et al. The two-minute walk test in persons with multiple sclerosis: correlations of cadence with free-living walking do not support ecological validity. *Int J Environ Res Public Health.* (2020) 17(23):1–11. doi: 10.3390/ijerph17239044
61. Ketelhut NB, Kindred JH, Pimentel RE, Hess AM, Tracy BL, Reiser RF, et al. Functional factors that are important correlates of physical activity in people with multiple sclerosis: a pilot study. *Disabil Rehabil.* (2018) 40(20):2416–23. doi: 10.1080/09638288.2017.1336647
62. Klaren RE, Hubbard EA, Motl RW, Pilutti LA, Wetter NC, Sutton BP. Objectively measured physical activity is associated with brain volumetric measurements in multiple sclerosis. *Behav Neurol.* (2015) 2015:482536. doi: 10.1155/2015/482536

63. Learmonth YC, Motl RW, Sandroff BM, Pula JH, Cadavid D. Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis. *BMC Neurol.* (2013) 13(1):37–37. doi: 10.1186/1471-2377-13-37
64. Learmonth YC, Kinnett-Hopkins D, Rice IM, Dysterheft JL, Motl RW. Accelerometer output and its association with energy expenditure during manual wheelchair propulsion. *Spinal Cord.* (2016) 54(2):110–4. doi: 10.1038/sc.2015.33
65. Moon Y, McGinnis RS, Seagers K, Motl RW, Sheth N, Wright JA, et al. Monitoring gait in multiple sclerosis with novel wearable motion sensors. *PLoS ONE.* (2017) 12(2):1–19. doi: 10.1371/journal.pone.0171346
66. Motl RW, Suh Y, Balantrapu S, Sandroff BM, Sosnoff JJ, Pula J, et al. Evidence for the different physiological significance of the 6- and 2-minute walk tests in multiple sclerosis. *BMC Neurol.* (2012) 12(6). doi: 10.1186/1471-2377-12-6
67. Motl RW, Pilutti L, Sandroff BM, Dlugonski D, Sosnoff JJ, Pula JH. Accelerometry as a behavioral intervention targeting symptoms and physical activity in multiple sclerosis. *Acta Neurol Scand.* (2013) 127(6):384–90. doi: 10.1111/ane.12036
68. Motl RW, Learmonth YC, Pilutti LA, Dlugonski D, Klaren R. Validity of minimal clinically important difference values for the multiple sclerosis walking scale-12? *Eur Neurol.* (2014) 71(3–4):196–202. doi: 10.1159/000356116
69. Aburub A, Khalil H, Al-Sharman A, Alomari M, Khabour O. The association between physical activity and sleep characteristics in people with multiple sclerosis. *Mult Scler Relat Disord.* (2017) 12(August 2016):29–33. doi: 10.1016/j.msard.2016.12.010
70. Motl RW, Dlugonski D, Pilutti LA, Klaren RE. Does the effect of a physical activity behavioral intervention vary by characteristics of people with multiple sclerosis? *Int J MS Care.* (2015) 17(2):65–72. doi: 10.7224/1537-2073.2014-016
71. Pilutti LA, Dlugonski D, Sandroff BM, Klaren R, Motl RW. Randomized controlled trial of a behavioral intervention targeting symptoms and physical activity in multiple sclerosis. *Mult Scler J.* (2014) 20(5):594–601. doi: 10.1177/1352458513503391
72. Sandroff BM, Pilutti LA, Dlugonski D, Motl RW. Physical activity and information processing speed in persons with multiple sclerosis: a prospective study. *Ment Health Phys Act.* (2013) 6(3):205–11. doi: 10.1016/j.mhpa.2013.08.001
73. Pau M, Porta M, Coghe G, Frau J, Lorefice L, Cocco E. Does multiple sclerosis differently impact physical activity in women and men? A quantitative study based on wearable accelerometers. *Int J Environ Res Public Health.* (2020) 17(23):1–13. doi: 10.3390/ijerph17238848
74. Pau M, Porta M, Coghe G, Cocco E. What gait features influence the amount and intensity of physical activity in people with multiple sclerosis? *Medicine (Baltimore).* (2021) 100(9):e24931. doi: 10.1097/MD.00000000000024931
75. Pau M, Leban B, Deidda M, Porta M, Coghe G, Cattaneo D, et al. Use of wrist-worn accelerometers to quantify bilateral upper limb activity and asymmetry under free-living conditions in people with multiple sclerosis. *Mult Scler Relat Disord.* (2021) 53:103081. doi: 10.1016/j.msard.2021.103081
76. Plow M, Finlayson M, Motl RW, Bethoux F. Randomized controlled trial of a teleconference fatigue management plus physical activity intervention in adults with multiple sclerosis: rationale and research protocol. *BMC Neurol.* (2012) 12:1–8. doi: 10.1186/1471-2377-12-122
77. Plow M, Finlayson M, Liu J, Motl RW, Bethoux F, Sattar A. Randomized controlled trial of a telephone-delivered physical activity and fatigue self-management interventions in adults with multiple sclerosis. *Arch Phys Med Rehabil.* (2019) 100(11):2006–14. doi: 10.1016/j.apmr.2019.04.022
78. Rice IM, Rice LA, Motl RW. Promoting physical activity through a manual wheelchair propulsion intervention in persons with multiple sclerosis. *Arch Phys Med Rehabil.* (2015) 96(10):1850–8. doi: 10.1016/j.apmr.2015.06.011
79. Ryan JM, Fortune J, Stennett A, Kilbride C, Anokye N, Victor C, et al. Changing physical activity behaviour for people with multiple sclerosis: protocol of a randomised controlled feasibility trial (iStep-MS). *BMJ Open.* (2017) 7(11):e018875. doi: 10.1136/bmjopen-2017-018875
80. Adamson BC, Learmonth YC, Kinnett-Hopkins D, Bohri M, Motl RW. Feasibility study design and methods for project GEMS: guidelines for exercise in multiple sclerosis. *Contemp Clin Trials.* (2016) 47:32–9. doi: 10.1016/j.cct.2015.12.002
81. Ryan JM, Fortune J, Stennett A, Kilbride C, Lavelle G, Hendrie W, et al. Safety, feasibility, acceptability and effects of a behaviour-change intervention to change physical activity behaviour among people with multiple sclerosis: results from the iStep-MS randomised controlled trial. *Mult Scler.* (2019) 26(14):1907–18. doi: 10.1177/1352458519886231
82. Sagawa Y, Watelain E, Moulin T, Decavel P. No effect of fampridine on real-life physical activity in people with multiple sclerosis. *Ann Phys Rehabil Med.* (2018) 61(2):105–7. doi: 10.1016/j.rehab.2017.10.006
83. Sandroff BM, Dlugonski D, Pilutti LA, Pula JH, Benedict RHB, Motl RW. Physical activity is associated with cognitive processing speed in persons with multiple sclerosis. *Mult Scler Relat Disord.* (2014) 3(1):123–8. doi: 10.1016/j.msard.2013.04.003
84. Sandroff BM, Motl RW, Kam JP, Pula JH. Accelerometer measured physical activity and the integrity of the anterior visual pathway in multiple sclerosis. *Mult Scler Relat Disord.* (2014) 3(1):117–22. doi: 10.1016/j.msard.2013.06.014
85. Sandroff BM, Motl RW, Pilutti LA, Learmonth YC, Ensari I, Dlugonski D, et al. Accuracy of StepWatch™ and ActiGraph accelerometers for measuring steps taken among persons with multiple sclerosis. *PLoS One.* (2014) 9(4):e93511. doi: 10.1371/journal.pone.0093511
86. Sandroff BM, Klaren RE, Motl RW. Relationships among physical inactivity, deconditioning, and walking impairment in persons with multiple sclerosis. *J Neurol Phys Ther.* (2015) 39(2):103–10. doi: 10.1097/NPT.0000000000000087
87. Sandroff BM, Motl RW, Sosnoff JJ, Pula JH. Further validation of the six-spot step test as a measure of ambulation in multiple sclerosis. *Gait Posture.* (2015) 41(1):222–7. doi: 10.1016/j.gaitpost.2014.10.011
88. Sandroff BM, Motl RW. Device-Measured physical activity and cognitive processing speed impairment in a large sample of persons with multiple sclerosis. *J Int Neuropsychol Soc.* (2020) 26(8):798–805. doi: 10.1017/S1355617720000284
89. Sasaki JE, Motl RW, McAuley E. Validity of the marshall sitting questionnaire in people with multiple sclerosis. *J Sports Sci.* (2019) 37(11):1250–6. doi: 10.1080/02640414.2018.1554614
90. Learmonth YC, Adamson BC, Kinnett-Hopkins D, Bohri M, Motl RW. Results of a feasibility randomised controlled study of the guidelines for exercise in multiple sclerosis project. *Contemp Clin Trials.* (2017) 54:84–97. doi: 10.1016/j.cct.2016.11.012
91. Sebastião E, Learmonth YC, Motl RW. Lower physical activity in persons with multiple sclerosis at increased fall risk. *Am J Phys Med Rehabil.* (2017) 96(5):357–61. doi: 10.1097/PHM.0000000000000581
92. Silveira SL, Motl RW. Do social cognitive theory constructs explain response heterogeneity with a physical activity behavioral intervention in multiple sclerosis? *Contemp Clin Trials Commun.* (2019) 15(November 2018):100366. doi: 10.1016/j.conctc.2019.100366
93. Sola-Valls N, Blanco Y, Sepulveda M, Llufrui S, Martínez-Lapiscina EH, La Puma D, et al. Walking function in clinical monitoring of multiple sclerosis by telemedicine. *J Neurol.* (2015) 262(7):1706–13. doi: 10.1007/s00415-015-7764-x
94. Sola-Valls N, Blanco Y, Sepúlveda M, Llufrui S, Martínez-Lapiscina EH, Zubizarreta I, et al. Combined walking outcome measures identify clinically meaningful response to prolonged-release fampridine. *Ther Adv Neurol Disord.* (2018) 11:1–10. doi: 10.1177/1756286418780007
95. Sosnoff JJ, Sandroff BM, Pula JH, Morrison SM, Motl RW. Falls and physical activity in persons with multiple sclerosis. *Mult Scler Int.* (2012) 2012:315620. doi: 10.1155/2012/315620
96. Stephens S, Berenbaum T, Finlayson M, Motl RW, Yeh EA. Youth with multiple sclerosis have low levels of fitness. *Mult Scler J.* (2021) 27(10):1597–605. doi: 10.1177/1352458520974360
97. Zhai Y, Nasser N, Pöttgen J, Gezelbash E, Heesen C, Stellmann J-PP. Smartphone accelerometry: a smart and reliable measurement of real-life physical activity in multiple sclerosis and healthy individuals. *Front Neurol.* (2020s) 11:688. doi: 10.3389/fneur.2020.00688
98. Abonie US, Hettinga FJ. Effect of a tailored activity pacing intervention on fatigue and physical activity behaviours in adults with multiple sclerosis. *Int J Environ Res Public Health.* (2021) 18(1):1–10. doi: 10.3390/ijerph18010017
99. Prakash RS, Janssen AL, Mangani H. Physical activity and sedentary behavior in MS: impact on dissociable correlates of cognitive functioning. *Mult Scler J.* (2017) 23(3, Suppl. 1):680–975. doi: 10.1519/JPT.0000000000000163
100. Arntzen EC, Straume B, Odeh F, Feys P, Normann B. Group-based, individualized, comprehensive core stability and balance intervention provides immediate and long-term improvements in walking in individuals with multiple sclerosis: a randomized controlled trial. *Physiother Res Int.* (2020) 25(1):e1798. doi: 10.1002/pri.1798
101. Baird JF, Cederberg KJ, Sikes EM, Silveira SL, Jeng B, Sasaki JE, et al. Physical activity and walking performance across the lifespan among adults with multiple sclerosis. *Mult Scler Relat Disord.* (2019) 35:36–41. doi: 10.1016/j.msard.2019.07.003
102. Balantrapu S, Sosnoff JJ, Pula JH, Sandroff BM, Motl RW. Leg spasticity and ambulation in multiple sclerosis. *Mult Scler Int.* (2014) 2014:649390. doi: 10.1155/2014/649390
103. Kozey-Keagle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc.* (2011) 43:1561–7. doi: 10.1249/MSS.0b013e31820ce174

104. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, inc. Accelerometer. *Med Sci Sports Exerc.* (1998) 30(5):777–81. doi: 10.1097/00005768-199805000-00021
105. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport.* (2011) 14(5):411–6. doi: 10.1016/j.jsams.2011.04.003
106. Agiovlasitis S, Sandroff BM, Motl RW. Step-rate cut-points for physical activity intensity in patients with multiple sclerosis: the effect of disability status. *J Neurol Sci.* (2016) 361:95–100. doi: 10.1016/j.jns.2015.12.027
107. Knaier R, Höchsmann C, Infanger D, Hinrichs T, Schmidt-Trucksäss A. Validation of automatic wear-time detection algorithms in a free-living setting of wrist-worn and hip-worn ActiGraph GT3X+. *BMC Public Health.* (2019) 19(1). doi: 10.1186/s12889-019-6568-9
108. Feito Y, Garner HR, Bassett DR. Evaluation of ActiGraph's low-frequency filter in laboratory and free-living environments. *Med Sci Sports Exerc.* (2015) 47(1):211–7. doi: 10.1249/MSS.0000000000000395
109. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Kowalsky RJ, Perera S, Brach JS. Accuracy of objective physical activity monitors in measuring steps in older adults. *Gerontol Geriatr Med.* (2018) 4:233372141878112. doi: 10.1177/2333721418781126
110. Veldhuijzen Van Zanten JJCS, Pilutti LA, Duda JL, Motl RW. Sedentary behaviour in people with multiple sclerosis: is it time to stand up against MS? *Mult Scler.* (2016) 22:1250–6. doi: 10.1177/1352458516644340
111. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. *Med Sci Sports Exerc.* (2014) 46(9):1816–24. doi: 10.1249/MSS.0000000000000289
112. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sport Exerc.* (2011) 43(8):1561–7. doi: 10.1249/MSS.0b013e31820ce174
113. Carr LJ, Mahar MT. Accuracy of intensity and inclinometer output of three activity monitors for identification of sedentary behavior and light-intensity activity. *J Obes.* (2012) 2012:1–9. doi: 10.1155/2012/460271
114. Johnston W, Judice PB, Molina Garcia P, Mühlen JM, Lykke Skovgaard E, Stang J, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *Br J Sports Med.* (2021) 55:780–93. doi: 10.1136/bjsports-2020-103147