

Original Paper

# Accuracy of the Apple Watch Series 4 and Fitbit Versa for Assessing Energy Expenditure and Heart Rate of Wheelchair Users During Treadmill Wheelchair Propulsion: Cross-sectional Study

Marius Lyng Danielsson<sup>1</sup>, MSc; Melanie Vergeer<sup>1,2</sup>, MSc; Guy Plasqui<sup>2</sup>, PhD; Julia Kathrin Baumgart<sup>1</sup>, PhD

<sup>1</sup>Centre for Elite Sports Research, Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

<sup>2</sup>Department of Nutrition and Movement Sciences, Maastricht University, Maastricht, Netherlands

**Corresponding Author:**

Marius Lyng Danielsson, MSc  
Centre for Elite Sports Research  
Department of Neuromedicine and Movement Science  
Norwegian University of Science and Technology  
Smistadvegen 11  
Trondheim, 7026  
Norway  
Phone: +47 47863154  
Email: [mlyngd@gmail.com](mailto:mlyngd@gmail.com)

## Abstract

**Background:** The Apple Watch (AW) Series 1 provides energy expenditure (EE) for wheelchair users but was found to be inaccurate with an error of approximately 30%, and the corresponding error for heart rate (HR) provided by the Fitbit Charge 2 was approximately 10% to 20%. Improved accuracy of estimated EE and HR is expected with newer editions of these smart watches (SWs).

**Objective:** This study aims to assess the accuracy of the AW Series 4 (wheelchair-specific setting) and the Fitbit Versa (treadmill running mode) for estimating EE and HR during wheelchair propulsion at different intensities.

**Methods:** Data from 20 manual wheelchair users (male: n=11, female: n=9; body mass: mean 75, SD 19 kg) and 20 people without a disability (male: n=11, female: n=9; body mass: mean 75, SD 11 kg) were included. Three 4-minute wheelchair propulsion stages at increasing speed were performed on 3 separate test days (0.5%, 2.5%, or 5% incline), while EE and HR were collected by criterion devices and the AW or Fitbit. The mean absolute percentage error (MAPE) was used to indicate the absolute agreement between the criterion device and SWs for EE and HR. Additionally, linear mixed model analyses assessed the effect of exercise intensity, sex, and group on the SW error. Interclass correlation coefficients were used to assess relative agreement between criterion devices and SWs.

**Results:** The AW underestimated EE with MAPEs of 29.2% (SD 22%) in wheelchair users and 30% (SD 12%) in people without a disability. The Fitbit overestimated EE with MAPEs of 73.9% (SD 7%) in wheelchair users and 44.7% (SD 38%) in people without a disability. Both SWs underestimated HR. The device error for EE and HR increased with intensity for both SWs (all comparisons:  $P < .001$ ), and the only significant difference between groups was found for HR in the AW (-5.27 beats/min for wheelchair users;  $P = .02$ ). There was a significant effect of sex on the estimation error in EE, with worse accuracy for the AW (-0.69 kcal/min;  $P < .001$ ) and better accuracy for the Fitbit (-2.08 kcal/min;  $P < .001$ ) in female participants. For HR, sex differences were found only for the AW, with a smaller error in female participants (5.23 beats/min;  $P = .02$ ). Interclass correlation coefficients showed poor to moderate relative agreement for both SWs apart from 2 stage-incline combinations (AW: 0.12-0.57 for EE and 0.11-0.86 for HR; Fitbit: 0.06-0.85 for EE and 0.03-0.29 for HR).

**Conclusions:** Neither the AW nor Fitbit were sufficiently accurate for estimating EE or HR during wheelchair propulsion. The AW underestimated EE and the Fitbit overestimated EE, and both SWs underestimated HR. Caution is hence required when using SWs as a tool for training intensity regulation and energy balance or imbalance in wheelchair users.

**KEYWORDS**

agreement; validity; accuracy; cross sectional; physiology; disability; disabled; upper-body exercise; upper body; exercise; physical activity; ergospirometer; fitness; vital; vitals; energy; expenditure; mHealth; wearable; wearables; mobile health; smartwatch; smartwatches; apple watch; fitbit; digital health; energy expenditure; heart rate; wheelchair; wheelchairs; fitness trackers; tracker; trackers

## Introduction

Wheelchair users are generally less active than people without a disability, which increases their risk of developing noncommunicable diseases such as cardiovascular disease, type 2 diabetes, and obesity [1-4]. In fact, the prevalence of obesity is 2.5 times greater in wheelchair users compared to people without a disability [5]. This is related to an energy intake that exceeds their energy expenditure (EE), which is approximately 5% to 40% lower in wheelchair users compared to people without a disability [6-8].

Total EE in people without a disability is approximately 1900 to 2900 kcal/day and comprised of 3 components: 60% to 75% attributed to resting EE (REE), 10% to diet-induced thermogenesis, and 15% to 30% to physical activity EE (PAEE) [9,10]. The lower total EE of wheelchair users is mainly related to reduced REE and PAEE [11-13]. Wheelchair users with a spinal cord injury (SCI) have a 14% to 27% lower REE due to reduced fat-free mass and sympathetic nervous system activity [11]. Furthermore, wheelchair users have lower PAEE due to a smaller amount of active muscle mass during upper-body exercise compared to what ambulatory people without a disability expend during walking or running [14,15]. Also, PAEE typically ranges between 6% and 36% of the total daily EE in wheelchair users with a SCI [16,17] and is the most modifiable of the 3 components. Therefore, PAEE may be particularly useful for obtaining a balance between EE and energy intake.

Criterion devices for measuring EE, such as direct or indirect calorimetry, are restricted to the laboratory setting and expensive to use. Therefore, more accessible devices that accurately estimate EE within the population of wheelchair users are needed. Smart watches (SWs) are widely used to provide feedback on estimated EE and monitor physical activity intensity (eg, through monitoring heart rate [HR]) [14,18-20]. If sufficiently accurate, the feedback provided by SWs may serve as a tool to counteract obesity and promote physical activity in wheelchair users.

Commonly used cutoffs for acceptable accuracy of parameters provided by wearable devices are  $\pm 10\%$  in free-living settings and  $\pm 3\%$  in standardized settings [18,21]. However, even in standardized settings, SWs often estimate HR and EE values outside of this range in both wheelchair users and people without a disability [18-22]. Additionally, the development of wheelchair

user-specific estimation algorithms for EE and HR is especially challenging because of the high heterogeneity and disability-related differences in physiological functioning in this population. Currently, the only study that evaluated the accuracy of estimated HR with a commercially available SW (Fitbit Charge 2 [Fitbit Inc]) found lower accuracy for wheelchair users (mean absolute percentage errors [MAPE] of approximately 10%-20% dependent on level of SCI) compared to people without a disability (approximately 8% MAPE) [23]. Furthermore, the sole study that assessed the accuracy of the estimated EE in the commercially available Apple Watch (AW) Series 1 (Apple Inc) with a wheelchair-specific setting reported a MAPE of 29% [24]. Notably, both studies report that the measurement error for estimating HR and EE increased with higher-intensity exercise [23,24]. While follow-up studies have not yet been conducted, one would expect companies to further improve the accuracy of their HR and EE estimation algorithms.

Therefore, the aim of this study was to assess the accuracy of the AW Series 4 (in the wheelchair-specific setting “outdoor push walking pace”) and the Fitbit Versa (in the treadmill running mode) for estimating EE and HR during wheelchair propulsion at different intensities. We decided to include both wheelchair users and a control group consisting of people without a disability to investigate if the wheelchair setting was specifically adjusted for wheelchair users.

## Methods

### Participants

A total of 20 wheelchair users and 20 people without a disability were included in the study. Both groups consisted of 11 male participants and 9 female participants and had similar demographic characteristics (Table 1). Participants were included if they were aged between 18 and 60 years and without injury or other health issues that could be aggravated by physical exertion. Included in the wheelchair user group were individuals that used a manual wheelchair as a main form of transport or were ambulatory wheelchair users. The wheelchair user group was comprised of individuals with SCI (n=11), spina bifida (n=2), and cerebral palsy (n=2). A total of 5 participants had other neurological, musculoskeletal, or joint impairments. Participants were recruited from sports associations and organizations for people with disabilities in Norway and social media.

**Table 1.** Participant characteristics.

Groups and sex	Age (years), mean (SD)	Body mass (kg), mean (SD)	Body height (cm), mean (SD)	BMI (kg/m <sup>2</sup> ), mean (SD)
<b>Combined</b>				
All	35.3 (11.8)	74.8 (15.2)	174.5 (10.9)	24.5 (4.1)
Male participants	36.3 (12.2)	81.1 (11.9)	181.9 (7.2)	24.4 (2.8)
Female participants	34.1 (11.5)	67.1 (15.6)	165.3 (7.1)	24.6 (5.3)
<b>Wheelchair users</b>				
All	37.4 (12.6)	74.5 (18.6)	172.5 (12.2)	24.9 (5.3)
Male participants	40.0 (12.9)	80.4 (14.3)	180.5 (8.5)	24.5 (3.3)
Female participants	34.1 (12.1)	67.2 (21.3)	162.7 (8.0)	25.3 (7.2)
<b>People without a disability</b>				
All	33.3 (10.8)	75.2 (11.4)	176.2 (9.9)	24.2 (2.4)
Male participants	32.6 (10.8)	81.9 (9.4)	183.5 (5.9)	24.3 (2.4)
Female participants	34.0 (11.5)	67.0 (7.8)	167.3 (5.3)	23.9 (2.6)

## Study Protocol

Three test days of wheelchair propulsion with different treadmill incline-speed combinations were conducted within 2 consecutive weeks. A minimum of 24 hours separated each test day, and sessions occurred at approximately the same time of day to account for diurnal variations. All test days started with a 5-minute warmup at a 0.5% incline at a self-chosen speed that corresponded to a rating of perceived exertion of 7-9 on the Borg scale [25]. Then, 3 standardized 4-minute stages were

performed at a predetermined incline for the day (either 0.5%, 2.5%, or 5%) with increasing speed across the stages. The order of the test days was counterbalanced. The speed at each incline was established through pilot testing and determined to be manageable for the participants (Table 2). Anthropometric data (age and sex) were collected before testing, and body mass and height were collected on the first test day. Participants were instructed to avoid high intensity training and alcohol consumption 24 hours before testing, avoid caffeine on the day of testing, and fast for at least 2 hours before testing.

**Table 2.** Overview of the standardized speeds for the 3 test days (0.5%, 2.5%, or 5% incline) for male participants (without tetraplegia) and female participants or male tetraplegic wheelchair users.

Test days and participants	Stage 1, speed (km/h)	Stage 2, speed (km/h)	Stage 3, speed (km/h)
<b>0.5% day</b>			
Male participants	4	6	8
Female participants or male participant with tetraplegia	3	5	7
<b>2.5% day</b>			
Male participants	3	4	5
Female participants or male participant with tetraplegia	2	3	4
<b>5% day</b>			
Male participants	2	3	4
Female participants or male participant with tetraplegia	1	2	3

## Equipment

Participants' body mass was measured using a Kistler force plate (Kistler 9286BA; Kistler Instruments AG) before the first test day. Body mass was determined for participants in the wheelchair user group while seated in their own wheelchair and obtained by subtracting the mass of the individual wheelchair (range 6.5-18.2 kg). All people without a disability were weighed while standing without any equipment. Participants wore a facemask (7450 V2 Series; Hans Rudolph Inc), which was connected to a Vyntus CPX ergospirometer with a mixing

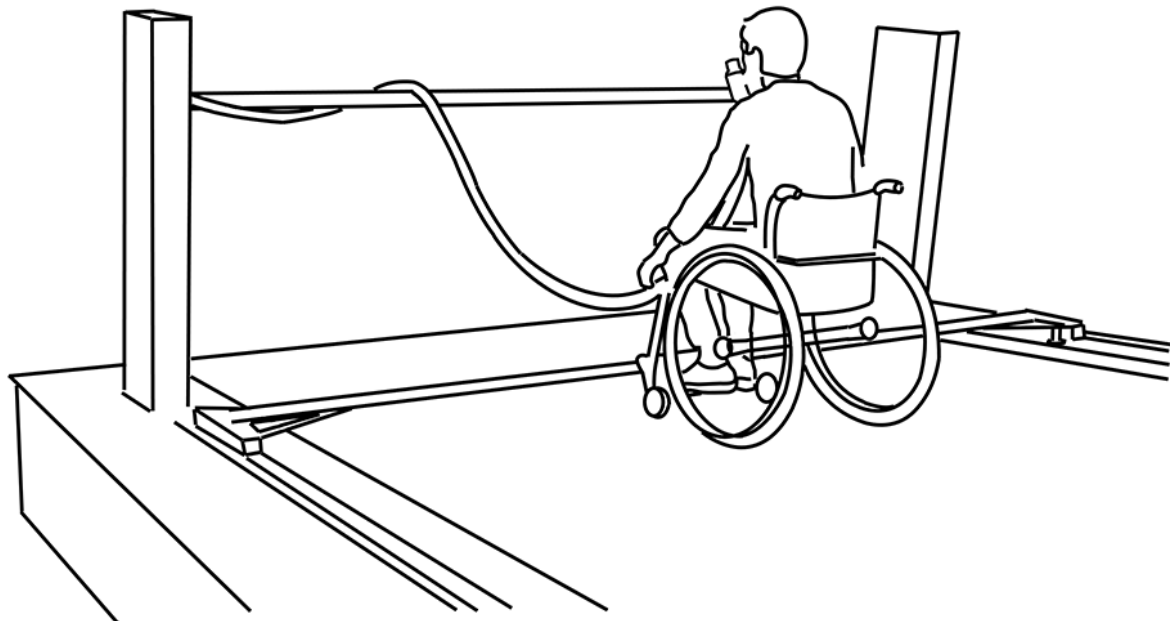
chamber (Vyaire, Medical GmbH) to measure gas exchange as 10-second averages, from which the criterion device EE was calculated. The Vyntus CPX was calibrated against a known gas mixture of 15% O<sub>2</sub> and 5% CO<sub>2</sub> before every test. Participants were fitted with a Polar HR monitor (version M400; Polar Electro Oy) and a Polar chest strap (version H10; Polar Electro Oy), which served as the criterion device for HR.

Participants wore 2 SWs on their nondominant wrist: an AW Series 4 software version OS 7.3.3 and a Fitbit Versa (2017) software version OS 5.0. The SWs tracked HR using photoplethysmography, and both SWs had a built-in

accelerometer and gyroscope. The SW placement (closest to the wrist) was counterbalanced. Participant characteristics were entered in the devices, and the activity settings of “outdoor push walking pace” (AW) and “treadmill running” (Fitbit) were used throughout testing. “Treadmill running” was chosen for the Fitbit in the absence of a wheelchair-specific setting. wheelchair users used their personal wheelchair, and people without a

disability used a standardized wheelchair (Küschall K-Series Attract; Invacare; mass 11.7kg). The wheelchairs were secured on a motorized 5×3 m treadmill (Forcelink Technology) with a mobile traverse bar attached to side rails (Figure 1). The side rails were equipped with safety stoppers to prevent participants from rolling off the back of the treadmill.

**Figure 1.** Test setup on the treadmill with the manual wheelchair attached to the traverse safety bar.



Blood lactate concentrations (mmol/L) were measured in a rested state and after every stage from a 20- $\mu$ l blood sample obtained from the participants earlobe [26]. Concentrations were analyzed using the Biosen C-Line Sport lactate measurement system (EKF Industrial Electronics). Rating of perceived exertion was obtained for muscular, respiratory, and total effort on the 6-20 Borg scale after each stage [25].

### Data Analysis

The criterion EE was calculated for every 10-second average  $VO_2$  and  $VCO_2$  values using the Weir formula [27]:

$$EE \text{ (kcal/min)} = 3.941 \times VO_2 \text{ (L/min)} + 1.106 \times VCO_2 \text{ (L/min)}$$

An average over the entire 4-minute period (as opposed to a steady-state average) was calculated for EE and HR for each stage of the criterion device. This was done since the SWs displayed an estimated average EE and HR for each entire stage. It was not possible to retrieve data with a higher time resolution from the SWs. The EE values provided by the criterion device and SWs were converted to kcal/min for comparison. Lastly, incomplete 4-minute stages were excluded from the analyses.

### Missing Data

In total, 20 of the 360 performed stages were incomplete and excluded from the analyses. Incomplete stages mostly occurred at the highest speeds and often at the 5% incline, which was due to a combination of an age- or disability-related lack of

physical fitness or upper-body strength. Additionally, equipment failure or human error contributed to the following missing data: incorrect activity setting (AW,  $n=2$ ), no HR recorded after activity (AW,  $n=32$ ), and a lost HR signal (criterion,  $n=1$ ).

### Statistical Analysis

Statistical analyses were conducted, and visualizations created in R Studio (version 4.2.1; R Core Team) [28]. Descriptive statistics were calculated for EE and HR for criterion devices, AW, and Fitbit and visualized with box plots using the R Studio package *ggplot2* (Multimedia Appendix 1).

### Absolute Agreement

The MAPE was used to establish the difference between criterion devices and SWs for both EE and HR during each stage:

$$MAPE = 1/n \times \sum [ (|C_p - SW_p|) / |C_p| ] \times 100$$

where criterion devices are represented by  $C$ , smart watches as  $SW$ , participants as  $p$ , and the total number of participants that completed the respective stage as  $n$ . In addition to the separate MAPEs visualized in a figure, an overall MAPE is provided in text, with the mean and SD being calculated across all participants and stages. The MAPE was categorized based on commonly used accuracy cutoffs for measuring EE, HR, and steps with wearable devices, with an acceptable error of  $\pm 10\%$  in free living settings and  $\pm 3\%$  in standardized settings [18,21]. Our categorization for the MAPE was therefore as follows: poor



(>20%), moderate (10.1%-20%), good (3.1%-10%) and excellent (0%-3%).

The results of linear mixed model analyses are presented in our main results, in addition to Bland-Altman plots in the [Multimedia Appendix 2](#). The main reason for choosing linear mixed model analyses as our main analyses was the repeated-measures design of our data collection with corresponding dependency in data [29], since all participants conducted several stages. The linear mixed model analyses were used to assess if there was a significant difference in EE and HR between criterion devices and SWs and to investigate the effect of group and sex as well as the increase in intensity on these device differences. Speed and incline were not adjusted for in the mixed model analyses, as we were interested in the estimation error of the SWs across all intensities and not within each speed-incline combination. As such, we also did not need to adjust for multiple comparisons. Participant ID was included as a random-intercept effect in these models to account for dependency in the data. An  $\alpha$  value of .05 was used to indicate statistical significance. There was no deviation of the residuals from normality (checked by visual inspection of Q-Q plots) and no violation of the assumption of homoscedasticity (checked by plotting the fitted values against the residuals; *plot\_model* function, R *sjPlot* package). The inclusion or exclusion of outliers did not change the results of the mixed model analyses. We therefore decided to include the analyses with outliers, as they represent the actual estimation errors of the AW.

### Relative Agreement

Interclass correlation coefficients (ICCs) were used to quantify relative agreement between criterion devices and SWs for both EE and HR. The ICCs were calculated using the 2-way random effects model with absolute agreement by using the *icc* function from the *irr* package. ICCs were categorized based on widely used cutoff points into poor (<0.5), moderate (0.5-0.75), good (0.75-0.9), and excellent relative agreement (>0.9) [30,31].

### Ethical Considerations

The data collection and processing were approved by the Norwegian Centre for Research Data (216680) and conducted in accordance with the Declaration of Helsinki. All participants were informed of the study purpose, design, potential risks, and the possibility to withdraw without penalty before signing the

consent form. Participation was voluntary and without financial compensation beyond insight into the individual's collected data. The collected data were deidentified.

## Results

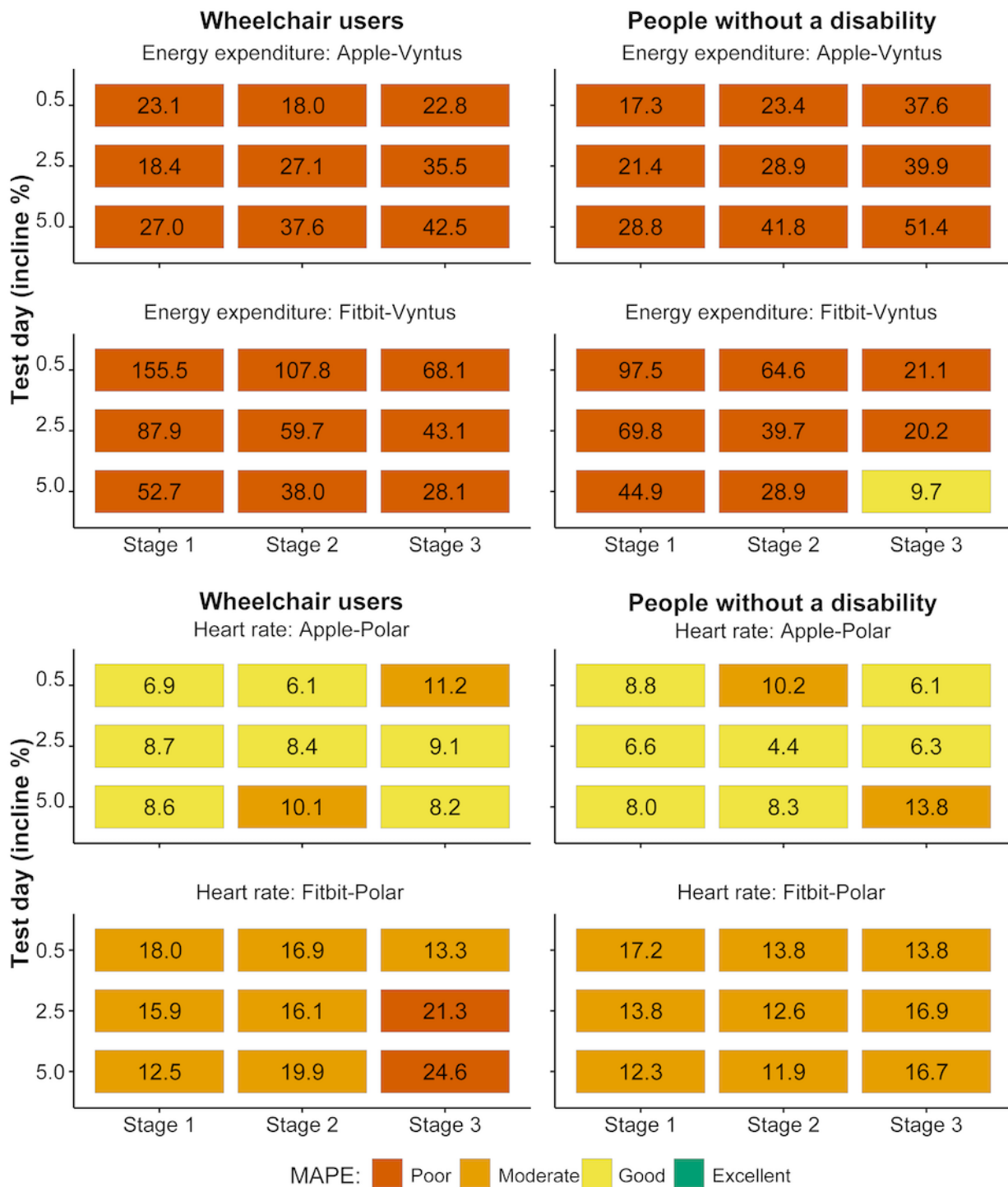
### Absolute Agreement

For the EE reported by the AW Series 4, the MAPE (with all inclines and stages combined) was 27.4% (SD 16.7%) in wheelchair users and 32.1% (SD 14.4%) in people without a disability. The EE provided by the Fitbit Versa had a MAPE of 73.9% (SD 57.2%) in wheelchair users and 44.7% (SD 37.8%) in people without a disability. Absolute agreement based on the MAPE for each incline-stage combination was mostly poor ([Figure 2](#)).

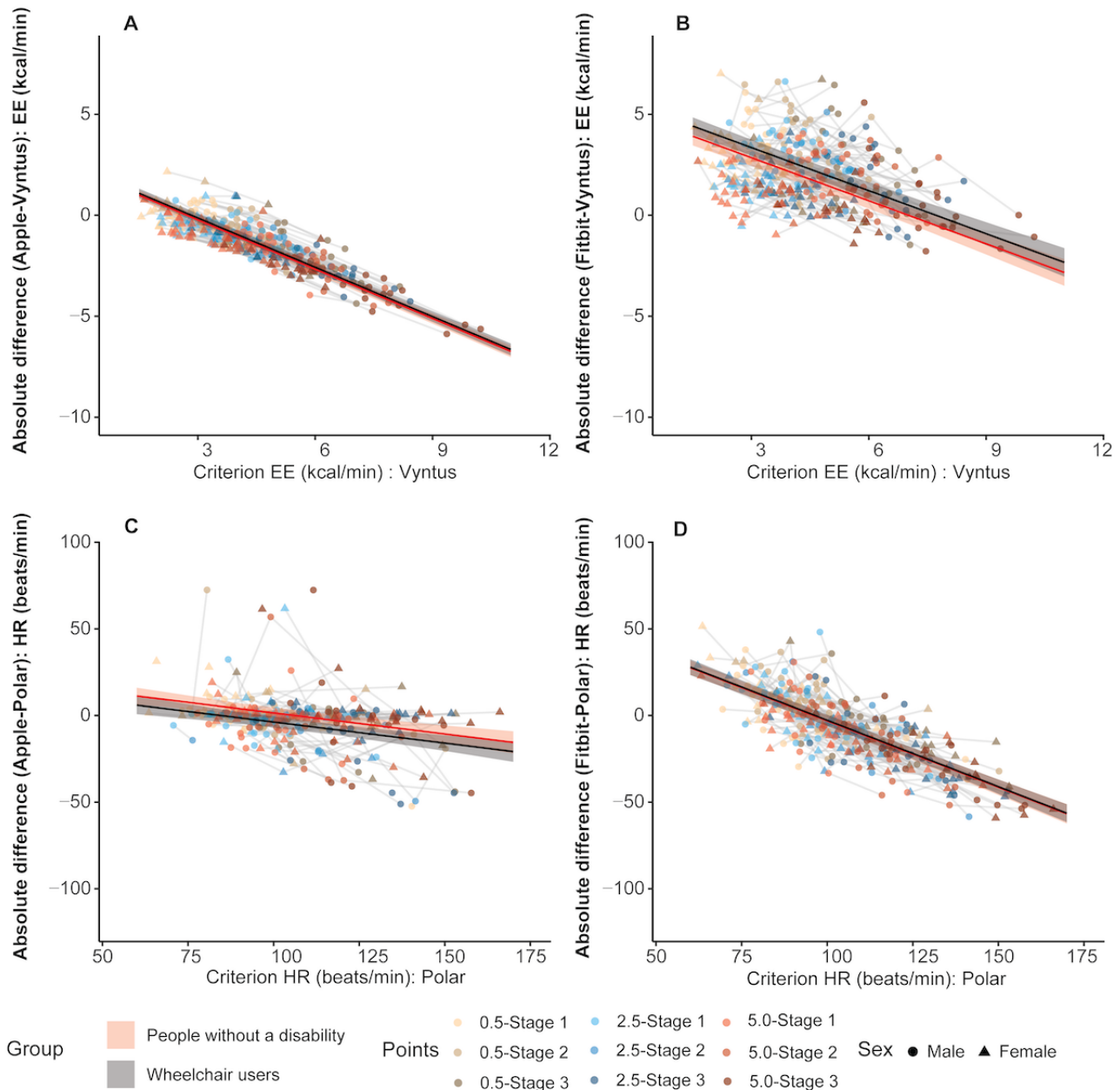
For HR, the MAPE with all stages and inclines combined was 8.5 (SD 10.4%) in wheelchair users and 8.1 (SD 13.6%) in people without a disability for the AW, and 17.4 (SD 12.4%) in wheelchair users and 14.3 (SD 10.7%) in people without a disability for the Fitbit. The absolute agreement for HR in each incline-stage combination was moderate-good for the AW and poor-moderate for the Fitbit ([Figure 2](#)).

The mixed model analyses indicated that EE was underestimated by the AW and overestimated by the Fitbit, while both SWs underestimated HR ([Figure 3](#)). Additionally, the differences between criterion and comparison devices increased negatively with higher exercise intensity for EE and HR (all comparisons,  $P<.001$ ; [Figure 3](#) and [Multimedia Appendix 2](#)). This led to reduced accuracy in the AW (larger underestimation) and improved accuracy in the Fitbit (lower overestimation) with increased intensity. EE and HR differences between SWs and criterion devices were not significantly different in wheelchair users compared to people without a disability in most comparisons ( $P>.06$ ), with the exception of the HR reported by the AW (-5.27 beats/min;  $P=.02$ ). Furthermore, for EE, the differences were significantly larger in female participants for the AW (-0.69 kcal/min;  $P<.001$ ) and in male participants for the Fitbit (-2.08 kcal/min;  $P<.001$ ). For HR, the only sex difference was found for the AW, with smaller differences between the AW and criterion device in female participants (5.23 beats/min;  $P=.02$ ).

**Figure 2.** Mean absolute percentage error (MAPE) between criterion devices and smart watches for energy expenditure and heart rate on all 3 test days (0.5%, 2.5%, and 5% incline) and stages with increasing speed. Values are presented separately for wheelchair users and people without a disability. MAPEs were categorized as poor (>20%), moderate (10.1%-20%), good (3.1%-10%), and excellent (0%-3%).



**Figure 3.** Regression lines (with shaded 95% CIs) separated by groups for the differences in energy expenditure (EE) and heart rate (HR) between smart watches (SWs) and criterion devices based on linear mixed model analyses: (A) EE: Apple-Vyntus; (B) EE: Fitbit-Vyntus; (C) HR: Apple-Polar; and (D): HR: Fitbit-Polar. The x-axis shows the criterion device values, while the y-axis shows the absolute difference between SWs and criterion devices. A regression line below zero indicates underestimation of SWs compared to criterion devices, while 1 above zero indicates overestimation.

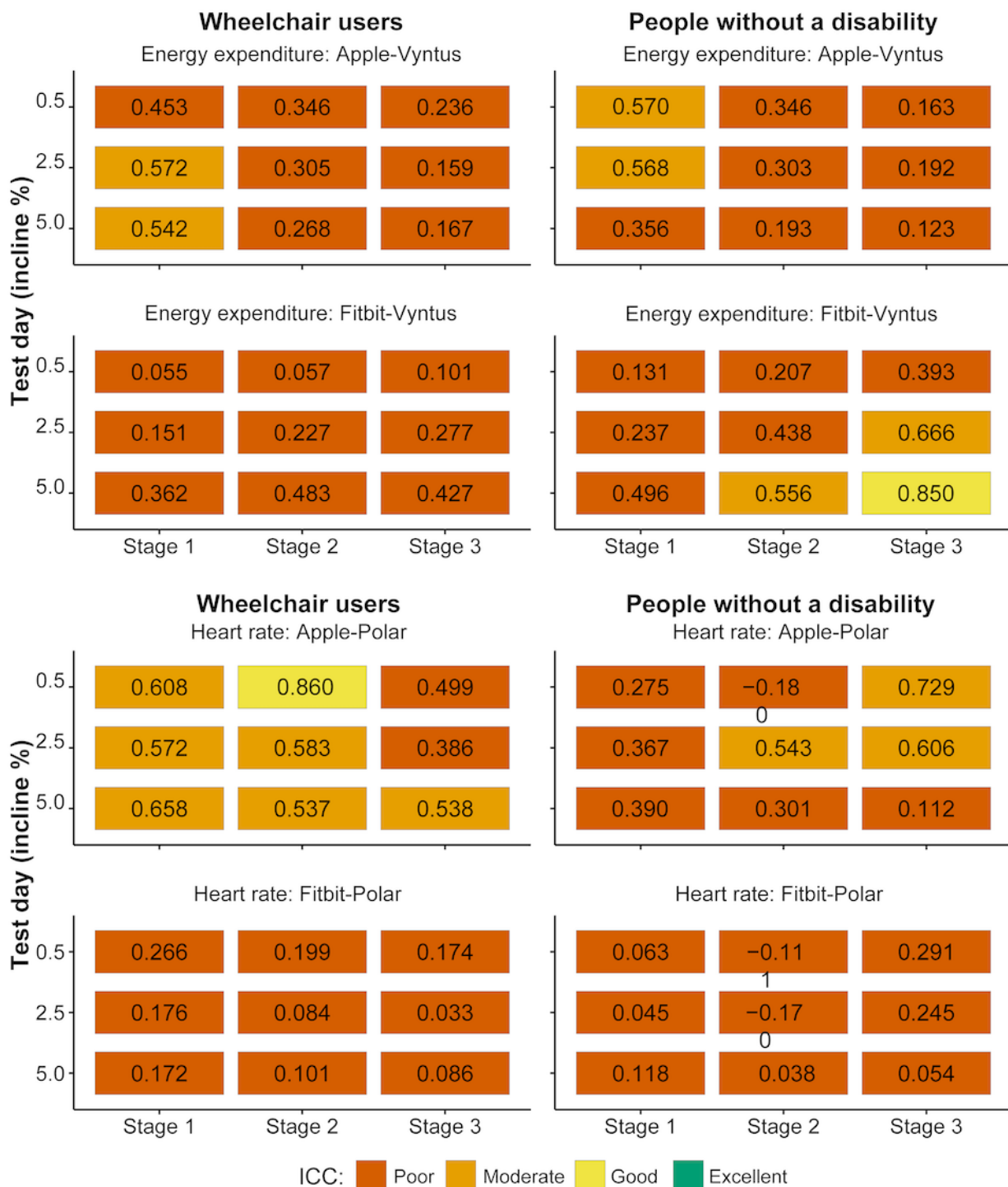


### Relative Agreement

Apart from 1 EE (Figure 4; a 5% incline for stage 3) and 1 HR ICC (Figure 4; a 0.5% incline for stage 2), all remaining ICCs indicate poor to moderate relative agreement between the

criterion devices and SWs (Figure 4). The ICCs for each incline-stage combination for the AW had a range from 0.12 to 0.57 for EE and from 0.11 to 0.86 for HR. For the Fitbit, the corresponding ranges were from 0.06 to 0.85 for EE and from 0.03 to 0.29 for HR.

**Figure 4.** Interclass correlation coefficients (ICCs) between criterion devices and smart watches (SWs) for wheelchair users and people without a disability on all 3 test days (0.5%, 2.5%, and 5% incline) and stages with increasing speed. ICCs were categorized as poor (<0.5), moderate (0.5-0.75), good (0.76-0.9), and excellent (>0.9) for energy expenditure and heart rate.



## Discussion

### Summary

The aim of this study was to assess the accuracy of the AW Series 4 and the Fitbit Versa for estimating EE and HR during wheelchair propulsion at different intensities. The AW underestimated EE and the Fitbit overestimated EE, suggesting

that neither of the SWs are accurate enough for estimating EE in wheelchair users. Furthermore, both the AW and Fitbit underestimated HR. Lastly, the differences in HR and EE between SWs and criterion devices increased with increasing intensity, and they mostly did not differ between groups.



## EE Findings

The MAPE of approximately 30% and poor to moderate ICCs of the AW Series 4 are similar to the ones reported by Moreno et al [24] for the AW Series 1. Since the participant characteristics were similar those in this study, there seems to be no improvement in the AW's EE estimation algorithms. The SW algorithms are proprietary technology, and it cannot therefore be determined why the AW underestimates EE for all participants. A plausible explanation is that Apple intentionally chose to report lower values for the sake of obesity prevention. Another possible explanation is that the data for developing the AW algorithms were collected from wheelchair users with lower training status or higher levels of SCI, who have a lower EE compared to the wheelchair users tested in this study. If the latter is the case, injury and fitness levels are important factors to consider when estimating the EE of wheelchair propulsion.

In this study, the AW was found to be equally inaccurate when estimating EE for wheelchair users and people without a disability, a finding that also aligns with Moreno et al [24]. Since the AW does not request user information on, for example, impairment levels, this finding indicates that the watch's software is not capable of identifying these types of individual characteristics from other factors such as movement patterns or HR responses. In contrast, sex could be inputted into the AW settings, and we found a significantly better estimated EE for male participants. Preliminary evidence [32] indicates that algorithms for wearable devices are developed mostly based on reference data from male participants and may therefore be less accurate for female participants.

The AW's underestimation of EE increased with intensity, which contradicts findings from previous AW studies [24,33]. While Pope et al [33] reported lower accuracy at moderate compared to low and high intensity during running, Moreno et al [24] reported consistent accuracy across wheelchair propulsion stages with an increasing stroke rate. The test protocol with the standardized stroke rate increases in Moreno et al [24] might more closely resemble the way the AW estimation algorithm works. Possibly, the AW uses the accelerometry data to determine how much exercise intensity, and thereby EE, has increased. In contrast, in this study, intensity increased with higher speeds at a given incline, with the steeper incline days being more physiologically taxing with a larger anaerobic contribution than the flatter incline days. Furthermore, it is possible that the AW EE estimation algorithms were developed mostly based on low-intensity data or without a physiological intensity measure (eg, HR) since wheelchair users spend most of their day at or below low-intensity exercise levels. In line with this, preliminary findings of our research group [34] indicate that estimation algorithms developed for wheelchair users perform less well on high-intensity data if they are only developed based on low- to moderate-intensity data.

The Fitbit did, in contrast to the AW and other SW studies on EE [24,33], show a systematic decrease in the error with higher intensities for both groups, which resulted in lower overestimations. This finding was most likely related to using the "treadmill running mode" in the absence of a wheelchair-specific setting. This setting leads the estimation

algorithm to expect a weight-bearing and leg-dependent activity with higher muscle activation. However, wheelchair propulsion is a non-weight-bearing activity that allows longer rest between cycles, especially at lower speeds and inclines. The reduction of the Fitbit error at higher intensities might therefore be a result of more active muscle mass, faster cycle rates, and longer cycle lengths with increased incline or speed. As such, the physiological effort of wheelchair propulsion may be more similar to running at higher intensities. Although the Fitbit displayed improved accuracy at higher intensities, the MAPEs were far greater than the  $\pm 3\%$  acceptable accuracy cutoff, and the Fitbit should therefore not be used to estimate EE in wheelchair users.

## HR Findings

Both SWs showed better accuracy for HR compared to EE. However, only the AW had a MAPE below the arbitrary cutoff of  $\pm 10\%$  that is commonly used for acceptable accuracy in free-living activities, with no values below the  $\pm 3\%$  cutoff for standardized settings [18,21]. Both SWs additionally showed reduced accuracy at the highest intensities, which is in agreement with previous findings of HR measured from wrist-worn devices during running [22]. Furthermore, wheelchair users were found to have a larger underestimation of HR (ie,  $-5$  beats/min) compared to people without a disability. While the reasons for this are somewhat unclear, it seems like this is due to more negative outliers in wheelchair users. Overall, the high MAPE variance and mostly poor to moderate (AW) and poor (Fitbit) ICCs indicate a high risk of individual inaccuracy when monitoring HR from these wrist-worn SWs during wheelchair propulsion.

## Methodological and Future Considerations

Two main factors need to be addressed for better EE and HR estimation algorithms in wrist-worn SWs: (1) the sensor hardware and (2) the sensor software (estimation algorithm). The sensor hardware of current wrist-worn technology is not capable of reporting precise or consistent HR signals during activity, which was partly highlighted by the missing data in this study. With regard to further improving the estimation algorithms, an assessment is needed on the impact and relative importance of factors such as personal characteristics (sex, age, body mass, training status, etc) or more wheelchair users-specific aspects such as type and level of impairment. These additions may increase estimation accuracy and reduce variation within the highly heterogeneous wheelchair user group, even without hardware improvements.

Furthermore, investigation into appropriate cutoffs for acceptable accuracy of EE and HR provided by SWs is needed. The commonly used cutoffs ( $\pm 3\%$  in standardized and  $\pm 10\%$  in free-living settings) are based on previous research assessing step count during walking [35-38] or pushes during wheelchair propulsion [39]. It may be plausible to establish higher cutoffs for EE or HR, especially when testing a heterogeneous wheelchair user group during an upper-body activity. However, a standardized low cutoff is essential to avoid differentiating acceptable accuracy between parameters in both controlled and free-living settings. Accurate estimates are also crucial for various populations, for example, to address an imbalance

between energy intake and expenditure in athletes trying to regulate nutrition for performance purposes or in wheelchair users attempting to prevent or counteract obesity. Therefore, it is likely the  $\pm 10\%$  error cutoff for free-living activities is too high.

Lastly, the effect of filtering done by the AW and Fitbit on the EE or HR data deserves mentioning. The AW removed many average HR values, which is likely attributed to a low number of data points. Comparatively, Fitbit reported data for all activities, although with lower accuracy. For now, we advise the use of HR belts for increased accuracy of the parameters investigated.

## Conclusion

Neither the wrist-worn AW nor Fitbit were sufficiently accurate for estimating EE or HR during wheelchair propulsion. The

AW underestimated EE while the Fitbit overestimated the EE across all incline-stage combinations. The underestimation of the AW increased and the overestimation of the Fitbit decreased with higher intensities, suggesting that neither watch sufficiently adjusts for the change in intensity. Additionally, both SWs underestimated HR. High MAPEs were found for both SWs and parameters (ie, EE and HR), in addition to the poor relative agreement indicated by low ICCs. Furthermore, neither the wheelchair-specific algorithm for estimating EE nor its ability to differentiate between wheelchair users and people without a disability have been improved for the AW Series 4 as compared to the previously investigated AW Series 1. Overall, our findings suggest that caution is required when using SWs as a tool for training intensity regulation and energy balance or imbalance in wheelchair users.

## Acknowledgments

Testing was performed with the laboratory facilities and equipment from NeXt Move, the Norwegian University of Science and Technology, the Faculty of Medicine at the Norwegian University of Science and Technology, and the Central Norway Regional Health Authority. We appreciate the work of everyone involved in “the digital wheelchair” project, in particular Cecilia Severin and Roy Mulder for contributions previously and during data collection, in addition to all the colleagues and master students assisting in the laboratory. This study was financed by strategic funding from the Faculty of Medicine and Health Science at the Norwegian University of Science and Technology, an innovation scholarship, and Discovery funding at the same university. The funding bodies had no role in the design of this study, its execution, analyses, interpretation of the data, or decision to submit results.

## Data Availability

The data sets generated during this study and corresponding analyses are available from the corresponding author on reasonable request. If used for further studies, the request includes a data application form and a data transfer agreement.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Descriptive box plots comparing absolute values for criterion energy expenditure from Vyntus and heart rate from Polar to comparison devices Apple Watch and Fitbit. Data were split between wheelchair users and people without a disability, and individual-case differences were added in (shaded gray).

[\[PPTX File , 775 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Bland-Altman plots visualizing absolute differences between criterion values and Apple or Fitbit values. The data were split between wheelchair users and people without disability and color coded for sex (male: blue circles, female: orange triangles). Furthermore, in accordance with Krouwer, criterion values were used on the x-axis rather than the average value of criterion and comparison device.

[\[PPTX File , 560 KB-Multimedia Appendix 2\]](#)

## References

1. Myers J, Lee M, Kiratli J. Cardiovascular disease in spinal cord injury: an overview of prevalence, risk, evaluation, and management. *Am J Phys Med Rehabil*. 2007;86(2):142-152. [doi: [10.1097/PHM.0b013e31802f0247](https://doi.org/10.1097/PHM.0b013e31802f0247)] [Medline: [17251696](https://pubmed.ncbi.nlm.nih.gov/17251696/)]
2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. *Can Med Assoc J*. 2006;174(6):801-809. [FREE Full text] [doi: [10.1503/cmaj.051351](https://doi.org/10.1503/cmaj.051351)] [Medline: [16534088](https://pubmed.ncbi.nlm.nih.gov/16534088/)]
3. Ginis KAM, Jørgensen S, Stapleton J. Exercise and sport for persons with spinal cord injury. *PM R*. 2012;4(11):894-900. [doi: [10.1016/j.pmrj.2012.08.006](https://doi.org/10.1016/j.pmrj.2012.08.006)] [Medline: [23174556](https://pubmed.ncbi.nlm.nih.gov/23174556/)]

4. Buchholz AC, Ginis KAM, Bray SR, Craven BC, Hicks AL, Hayes KC, et al. Greater daily leisure time physical activity is associated with lower chronic disease risk in adults with spinal cord injury. *Appl Physiol Nutr Metab*. 2009;34(4):640-647. [doi: [10.1139/H09-050](https://doi.org/10.1139/H09-050)] [Medline: [19767799](https://pubmed.ncbi.nlm.nih.gov/19767799/)]
5. Weil E, Wachterman M, McCarthy EP, Davis RB, O'Day B, Iezzoni LI, et al. Obesity among adults with disabling conditions. *JAMA*. 2002;288(10):1265-1268. [FREE Full text] [doi: [10.1001/jama.288.10.1265](https://doi.org/10.1001/jama.288.10.1265)] [Medline: [12215134](https://pubmed.ncbi.nlm.nih.gov/12215134/)]
6. Farkas GJ, Sneij A, Gater DR. Energy expenditure following spinal cord injury: a delicate balance. *Top Spinal Cord Inj Rehabil*. 2021;27(1):92-99. [FREE Full text] [doi: [10.46292/sci20-00030](https://doi.org/10.46292/sci20-00030)] [Medline: [33814887](https://pubmed.ncbi.nlm.nih.gov/33814887/)]
7. Shea JR, Shay BL, Leiter J, Cowley KC. Energy expenditure as a function of activity level after spinal cord injury: the need for tetraplegia-specific energy balance guidelines. *Front Physiol*. 2018;9:1286. [FREE Full text] [doi: [10.3389/fphys.2018.01286](https://doi.org/10.3389/fphys.2018.01286)] [Medline: [30283348](https://pubmed.ncbi.nlm.nih.gov/30283348/)]
8. Johnson RK, Hildreth HG, Contompasis SH, Goran MI. Total energy expenditure in adults with cerebral palsy as assessed by doubly labeled water. *J Am Diet Assoc*. 1997;97(9):966-970. [doi: [10.1016/s0002-8223\(97\)00233-2](https://doi.org/10.1016/s0002-8223(97)00233-2)] [Medline: [9284872](https://pubmed.ncbi.nlm.nih.gov/9284872/)]
9. National Research Council Subcommittee on the Tenth Edition of the Recommended Dietary A. The national academies collection: reports funded by national institutes of health. In: *Recommended Dietary Allowances: 10th Edition*. Washington (DC). National Academies Press (US); 1989;10.
10. Poehlman ET. A review: exercise and its influence on resting energy metabolism in man. *Med Sci Sports Exerc*. 1989;21(5):515-525. [Medline: [2691813](https://pubmed.ncbi.nlm.nih.gov/2691813/)]
11. Buchholz AC, Pencharz PB. Energy expenditure in chronic spinal cord injury. *Curr Opin Clin Nutr Metab Care*. 2004;7(6):635-639. [doi: [10.1097/00075197-200411000-00008](https://doi.org/10.1097/00075197-200411000-00008)] [Medline: [15534431](https://pubmed.ncbi.nlm.nih.gov/15534431/)]
12. Buchholz AC, McGillivray CF, Pencharz PB. Differences in resting metabolic rate between paraplegic and able-bodied subjects are explained by differences in body composition. *Am J Clin Nutr*. 2003;77(2):371-378. [FREE Full text] [doi: [10.1093/ajcn/77.2.371](https://doi.org/10.1093/ajcn/77.2.371)] [Medline: [12540396](https://pubmed.ncbi.nlm.nih.gov/12540396/)]
13. de Hollander EL, Proper KI. Physical activity levels of adults with various physical disabilities. *Prev Med Rep*. 2018;10:370-376. [FREE Full text] [doi: [10.1016/j.pmedr.2018.04.017](https://doi.org/10.1016/j.pmedr.2018.04.017)] [Medline: [29755934](https://pubmed.ncbi.nlm.nih.gov/29755934/)]
14. Nightingale TE, Rouse PC, Thompson D, Bilzon JLJ. Measurement of physical activity and energy expenditure in wheelchair users: methods, considerations and future directions. *Sports Med Open*. 2017;3(1):10. [FREE Full text] [doi: [10.1186/s40798-017-0077-0](https://doi.org/10.1186/s40798-017-0077-0)] [Medline: [28251597](https://pubmed.ncbi.nlm.nih.gov/28251597/)]
15. Glaser RM, Sawka MN, Wilde SW, Woodrow BK, Suryaprasad AG. Energy cost and cardiopulmonary responses for wheelchair locomotion and walking on tile and on carpet. *Paraplegia*. 1981;19(4):220-226. [FREE Full text] [doi: [10.1038/sc.1981.45](https://doi.org/10.1038/sc.1981.45)] [Medline: [7290731](https://pubmed.ncbi.nlm.nih.gov/7290731/)]
16. Nightingale TE, Williams S, Thompson D, Bilzon JLJ. Energy balance components in persons with paraplegia: daily variation and appropriate measurement duration. *Int J Behav Nutr Phys Act*. 2017;14(1):132. [FREE Full text] [doi: [10.1186/s12966-017-0590-z](https://doi.org/10.1186/s12966-017-0590-z)] [Medline: [28950900](https://pubmed.ncbi.nlm.nih.gov/28950900/)]
17. Westerterp KR. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Front Physiol*. 2013;4:90. [FREE Full text] [doi: [10.3389/fphys.2013.00090](https://doi.org/10.3389/fphys.2013.00090)] [Medline: [23637685](https://pubmed.ncbi.nlm.nih.gov/23637685/)]
18. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR Mhealth Uhealth*. 2020;8(9):e18694. [FREE Full text] [doi: [10.2196/18694](https://doi.org/10.2196/18694)] [Medline: [32897239](https://pubmed.ncbi.nlm.nih.gov/32897239/)]
19. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PLoS One*. 2017;12(2):e0171720. [FREE Full text] [doi: [10.1371/journal.pone.0171720](https://doi.org/10.1371/journal.pone.0171720)] [Medline: [28234979](https://pubmed.ncbi.nlm.nih.gov/28234979/)]
20. O'Driscoll R, Turicchi J, Hopkins M, Gibbons C, Larsen SC, Palmeira AL, et al. The validity of two widely used commercial and research-grade activity monitors, during resting, household and activity behaviours. *Health Technol*. 2019;10(3):637-648. [FREE Full text] [doi: [10.1007/s12553-019-00392-7](https://doi.org/10.1007/s12553-019-00392-7)]
21. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR Mhealth Uhealth*. 2018;6(8):e10527. [FREE Full text] [doi: [10.2196/10527](https://doi.org/10.2196/10527)] [Medline: [30093371](https://pubmed.ncbi.nlm.nih.gov/30093371/)]
22. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *J Sports Sci*. 2019;37(12):1411-1419. [doi: [10.1080/02640414.2018.1560644](https://doi.org/10.1080/02640414.2018.1560644)] [Medline: [30657025](https://pubmed.ncbi.nlm.nih.gov/30657025/)]
23. Hoevenaars D, Yocarini IE, Paraschiakos S, Holla JFM, de Groot S, Kraaij W, et al. Accuracy of heart rate measurement by the Fitbit Charge 2 during wheelchair activities in people with spinal cord injury: instrument validation study. *JMIR Rehabil Assist Technol*. 2022;9(1):e27637. [FREE Full text] [doi: [10.2196/27637](https://doi.org/10.2196/27637)] [Medline: [35044306](https://pubmed.ncbi.nlm.nih.gov/35044306/)]
24. Moreno D, Glasheen E, Domingo A, Panaligan VB, Penafior T, Rioveros A, et al. Validity of caloric expenditure measured from a wheelchair user smartwatch. *Int J Sports Med*. 2020;41(8):505-511. [doi: [10.1055/a-1088-5629](https://doi.org/10.1055/a-1088-5629)] [Medline: [32176933](https://pubmed.ncbi.nlm.nih.gov/32176933/)]
25. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14(5):377-381. [Medline: [7154893](https://pubmed.ncbi.nlm.nih.gov/7154893/)]
26. Moran P, Prichard JG, Ansley L, Howatson G. The influence of blood lactate sample site on exercise prescription. *J Strength Cond Res*. 2012;26(2):563-567. [FREE Full text] [doi: [10.1519/JSC.0b013e318225f395](https://doi.org/10.1519/JSC.0b013e318225f395)] [Medline: [22240552](https://pubmed.ncbi.nlm.nih.gov/22240552/)]

27. de V Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol.* 1949;109(1-2):1-9. [FREE Full text] [doi: [10.1113/jphysiol.1949.sp004363](https://doi.org/10.1113/jphysiol.1949.sp004363)] [Medline: [15394301](https://pubmed.ncbi.nlm.nih.gov/15394301/)]
28. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.; 2013. URL: <https://www.r-project.org/> [accessed 2024-04-03]
29. Parker RA, Scott C, Inácio V, Stevens NT. Using multiple agreement methods for continuous repeated measures data: a tutorial for practitioners. *BMC Med Res Methodol.* 2020;20(1):154. [FREE Full text] [doi: [10.1186/s12874-020-01022-x](https://doi.org/10.1186/s12874-020-01022-x)] [Medline: [32532218](https://pubmed.ncbi.nlm.nih.gov/32532218/)]
30. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155-163. [FREE Full text] [doi: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)] [Medline: [27330520](https://pubmed.ncbi.nlm.nih.gov/27330520/)]
31. St-Onge M, Mignault D, Allison DB, Rabasa-Lhoret R. Evaluation of a portable device to measure daily energy expenditure in free-living adults. *Am J Clin Nutr.* 2007;85(3):742-749. [FREE Full text] [doi: [10.1093/ajcn/85.3.742](https://doi.org/10.1093/ajcn/85.3.742)] [Medline: [17344495](https://pubmed.ncbi.nlm.nih.gov/17344495/)]
32. Musa SB, Ellis R, Chafe B, Sturrock SL, Maher RA, Cullen K, et al. Wearable device validity in measuring steps, energy expenditure, and heart rate across age, gender, and body mass index: data analysis from a systematic review. *J Phys Act Health.* 2023;20(2):100-105. [doi: [10.1123/jpah.2022-0160](https://doi.org/10.1123/jpah.2022-0160)] [Medline: [36535270](https://pubmed.ncbi.nlm.nih.gov/36535270/)]
33. Pope ZC, Zeng N, Li X, Liu W, Gao Z. Accuracy of commercially available smartwatches in assessing energy expenditure during rest and exercise. *J Meas Phys Behav.* 2019;2(2):73-81. [doi: [10.1123/jmpb.2018-0037](https://doi.org/10.1123/jmpb.2018-0037)]
34. Doshmanziari R, Aandahl HS, Danielsson ML, Baumgart JK, Varagnolo D. Experiment design considerations for estimating energy expenditure during wheelchair propulsion. *IFAC-PapersOnLine.* 2023;56(2):6504-6509. [FREE Full text] [doi: [10.1016/j.ifacol.2023.10.297](https://doi.org/10.1016/j.ifacol.2023.10.297)]
35. Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A. Evaluation of quality of commercial pedometers. *Can J Public Health.* 2006;97 Suppl 1(Suppl 1):S10-S15, S10-S16. [FREE Full text] [doi: [10.1007/BF03405359](https://doi.org/10.1007/BF03405359)] [Medline: [16676833](https://pubmed.ncbi.nlm.nih.gov/16676833/)]
36. Schneider PL, Crouter SE, Bassett DR. Pedometer measures of free-living physical activity: comparison of 13 models. *Med Sci Sports Exerc.* 2004;36(2):331-335. [FREE Full text] [doi: [10.1249/01.MSS.0000113486.60548.E9](https://doi.org/10.1249/01.MSS.0000113486.60548.E9)] [Medline: [14767259](https://pubmed.ncbi.nlm.nih.gov/14767259/)]
37. Schneider PL, Crouter SE, Lukajic O, Bassett DR. Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. *Med Sci Sports Exerc.* 2003;35(10):1779-1784. [FREE Full text] [doi: [10.1249/01.MSS.0000089342.96098.C4](https://doi.org/10.1249/01.MSS.0000089342.96098.C4)] [Medline: [14523320](https://pubmed.ncbi.nlm.nih.gov/14523320/)]
38. Bassett DR, Mahar MT, Rowe DA, Morrow JR. Walking and measurement. *Med Sci Sports Exerc.* 2008;40(7 Suppl):S529-S536. [FREE Full text] [doi: [10.1249/MSS.0b013e31817c699c](https://doi.org/10.1249/MSS.0b013e31817c699c)] [Medline: [18562970](https://pubmed.ncbi.nlm.nih.gov/18562970/)]
39. Glasheen E, Domingo A, Kressler J. Accuracy of Apple Watch fitness tracker for wheelchair use varies according to movement frequency and task. *Ann Phys Rehabil Med.* 2021;64(1):101382. [FREE Full text] [doi: [10.1016/j.rehab.2020.03.007](https://doi.org/10.1016/j.rehab.2020.03.007)] [Medline: [32335302](https://pubmed.ncbi.nlm.nih.gov/32335302/)]

## Abbreviations

- AW:** Apple Watch
- EE:** energy expenditure
- HR:** heart rate
- ICC:** interclass correlation coefficient
- MAPE:** mean absolute percentage error
- PAEE:** physical activity energy expenditure
- REE:** resting energy expenditure
- SCI:** spinal cord injury
- SW:** smart watch

*Edited by A Mavragani; submitted 05.09.23; peer-reviewed by H Aghayan Golkashani; comments to author 28.01.24; revised version received 19.02.24; accepted 26.02.24; published 07.05.24*

### *Please cite as:*

*Danielsson ML, Vergeer M, Plasqui G, Baumgart JK*

*Accuracy of the Apple Watch Series 4 and Fitbit Versa for Assessing Energy Expenditure and Heart Rate of Wheelchair Users During Treadmill Wheelchair Propulsion: Cross-sectional Study*

*JMIR Form Res 2024;8:e52312*

*URL: <https://formative.jmir.org/2024/1/e52312>*

*doi: [10.2196/52312](https://doi.org/10.2196/52312)*

*PMID:*

©Marius Lyng Danielsson, Melanie Vergeer, Guy Plasqui, Julia Kathrin Baumgart. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 07.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.