

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Le, Shenglong; Wang, Xiuqiang; Zhang, Tao; Lei, Si Man; Cheng, Sulin; Yao, Wu; Schumann, Moritz

Title: Validity of three smartwatches in estimating energy expenditure during outdoor walking and running

Year: 2022

Version: Published version

Copyright: © 2022 Le, Wang, Zhang, Lei, Cheng, Yao and Schumann.

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Le, S., Wang, X., Zhang, T., Lei, S. M., Cheng, S., Yao, W., & Schumann, M. (2022). Validity of three smartwatches in estimating energy expenditure during outdoor walking and running. *Frontiers in Physiology*, 13, Article 995575. <https://doi.org/10.3389/fphys.2022.995575>



OPEN ACCESS

EDITED BY

Hun-Young Park,
Konkuk University, South Korea

REVIEWED BY

Sang-Seok Nam,
Kukkiwon, South Korea
Sung-Woo Kim,
Konkuk University, South Korea

*CORRESPONDENCE

Wu Yao,
yaowu@sjtu.edu.cn
Moritz Schumann,
m.schumann@dshs-koeln.de

[†]These authors contributed equally to
this work and share last authorship

SPECIALTY SECTION

This article was submitted
to Exercise Physiology,
a section of the journal
Frontiers in Physiology

RECEIVED 16 July 2022

ACCEPTED 17 August 2022

PUBLISHED 26 September 2022

CITATION

Le S, Wang X, Zhang T, Lei SM, Cheng S,
Yao W and Schumann M (2022), Validity
of three smartwatches in estimating
energy expenditure during outdoor
walking and running.
Front. Physiol. 13:995575.
doi: 10.3389/fphys.2022.995575

COPYRIGHT

© 2022 Le, Wang, Zhang, Lei, Cheng,
Yao and Schumann. 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.

Validity of three smartwatches in estimating energy expenditure during outdoor walking and running

Shenglong Le^{1,2,3,4}, Xiuqiang Wang^{1,3}, Tao Zhang^{1,3}, Si Man Lei^{1,5},
Sulin Cheng^{1,3,4,6}, Wu Yao^{6*†} and Moritz Schumann^{7*†}

¹Exercise Translational Medicine Center, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China, ²Department of Physical Therapy, Taihe Hospital, Hubei University of Medicine, Shiyan, China, ³Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China, ⁴Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland, ⁵Faculty of Education, University of Macao, Macao, China, ⁶Physical Education Department, Shanghai Jiao Tong University, Shanghai, China, ⁷Department of Molecular and Cellular Sport Medicine, German Sport University, Cologne, Germany

Commercially wrist-worn devices often present inaccurate estimations of energy expenditure (EE), with large between-device differences. We aimed to assess the validity of the Apple Watch Series 6 (AW), Garmin FENIX 6 (GF) and Huawei Watch GT 2e (HW) in estimating EE during outdoor walking and running. Twenty young normal-weight Chinese adults concurrently wore three index devices randomly positioned at both wrists during walking at 6 km/h and running at 10 km/h for 2 km on a 400-meter track. As a criterion, EE was assessed by indirect calorimetry (COSMED K5). For walking, EE from AW and GF was significantly higher than that obtained by the K5 ($p < 0.001$ and 0.002 , respectively), but not for HW ($p = 0.491$). The mean absolute percentage error (MAPE) was 19.8% for AW, 32.0% for GF, and 9.9% for HW, respectively. The limits of agreement (LoA) were 44.1, 150.1 and 48.6 kcal for AW, GF, and HW respectively. The intraclass correlation coefficient (ICC) was 0.821, 0.216 and 0.760 for AW, GF, and HW, respectively. For running, EE from AW and GF were significantly higher than the K5 ($p < 0.001$ and 0.001 , respectively), but not for HW ($p = 0.946$). The MAPE was 24.4%, 21.8% and 11.9% for AW, GF and HW, respectively. LoA were 62.8, 89.4 and 65.6 kcal for AW, GF and HW, respectively. The ICC was 0.741, 0.594, and 0.698 for AW, GF and HW, respectively. The results indicate that the tested smartwatches show a moderate validity in EE estimations for outdoor walking and running.

KEYWORDS

wearable devices, validation, accuracy, physical activity, health monitoring

Introduction

Wearable technology has been ranked as the top fitness trend for several years (Thompson, 2021) and in 2021 the worldwide shipments of wearable devices reached 533.6 million units (Laricchia, 2022). Among these, wrist-worn devices are most common and are capable of monitoring a large variety of vital parameters including steps taken, distance travelled, heart rate and energy expenditure (EE) with a variety of sensors. Given their convenience, the data provided by wearable devices are often used to monitor and/or modify health behaviors both for self-monitoring as well as in healthcare and research settings (Lyons et al., 2014). The use of the devices is typically encouraged by marketing claims set out by the manufacturers, while actual data on the validity of the devices is typically lacking (Evenson et al., 2015).

Estimating EE by wrist-worn wearable devices seems convenient, considering the linear relationship of heart rate and/or physical activity with gaseous exchange. Consequently, numerous validation studies on EE estimations have been conducted, but provided heterogeneous results (O'driscoll et al., 2020; Argent et al., 2022). Overall, it appears that commercially available wrist- or arm-worn devices did not show a sufficient accuracy, while large between-device differences were observed (O'driscoll et al., 2020).

The differences in the accuracy of different devices may be dependent on several factors. Most importantly, the accuracy of the input parameters, such as the accuracy of heart rate from photoplethysmography (PPG) sensors affects EE estimations (O'driscoll et al., 2020; Argent et al., 2022). In this context, it appears that the accuracy of EE estimation may vary with the type and intensity of activities (O'driscoll et al., 2020). While this may be related to motion affecting the accuracy of the PPG signal, the observed error for different types of activities may also be related to the algorithms that likely do not take physical activity type or bodily posture into account (Schneller et al., 2015). Several studies found that algorithm adjustments may indeed improve the validity of the EE estimation during exercise (Jakicic et al., 2004; Van Hoye et al., 2015). Furthermore, addition of heart rate or heat sensors to accelerometer can improve the accuracy of EE estimations compared to accelerometry alone (O'driscoll et al., 2020; Kinnunen et al., 2019). Collectively, it appears that further technological advancements and revised algorithms may improve the quality of EE estimations. Thus, continuous validations of devices that are newly introduced to the market are indispensable.

For the thorough validation of EE estimations, a number of factors should be considered. Among the most important variables appears to be the intended use of the device, i.e. validating EE estimations in settings that represent the actual use by consumers (Argent et al., 2022). This is because specific sensors may be used for specific activities. In this regard, outdoor activities may be supported by GPS signal while indoor activities may solely rely on accelerometry, likely compromising the accuracy (Charlot et al., 2014). Therefore, this study aims to examine the accuracy of the EE

estimations of three new generations of smartwatches of the leading manufacturers, namely the Apple Watch Series 6 (AW), Garmin Fenix 6 (GF) and Huawei GT 2e (HW) during outdoor walking and running in young Chinese adults.

Materials and methods

Participants

Twenty healthy Chinese subjects (10 males and 10 females) were recruited from the local university campus (Table 1). On the bases of the paired sample *t*-test, a post-hoc analysis was performed to estimate if sample size is sufficient with proper effect size from three devices under both walking and running conditions by G*Power version 3.1.9.6 (Franz Faul, University Kiel, Germany) using the mean and standard deviations between criterion device (K5) and test device (Passler et al., 2019). For AW, GF and HW, the statistical power is 1.000, 0.929 and 0.105 for the outdoor walking, respectively, while 0.999, 0.942 and 0.051 for the outdoor running respectively. Participants were screened for inclusion criteria using the lifestyle and disease questionnaire and the ACSM Medical History, Signs and Symptoms, and Risk Factors for Risk Stratification (Jonas, 2009). Inclusion criteria included healthy young adults (aged 18–30 years) with a normal weight (body mass index 18.5–25 kg/m²) and self-reported regular recreational exercise (≥ 3 weekly exercise sessions). All participants were informed about the study procedures and provided written informed consent prior to the testing. The study was conducted in accordance with the declaration of Helsinki and approved by the Institutional Review Board for Human Research Protections of Shanghai Jiao Tong University (registration number B2020024I). Criterion measure assessment.

As the criterion, the Cosmed K5 (K5) system (Cosmed, Italy) was used. K5 is a portable gas analysis system using Breath-by-Breath technology. This technology enables a precise and accurate determination of the individual VE, VO₂ and VCO₂ for a wide range of metabolic rates (Deblois et al., 2021). The estimation of EE is based on the ratio of inhaled oxygen to exhaled carbon dioxide. Respiratory gases of each breath are analyzed. Before each measurement, the device was warmed up for a minimum of 15 min and calibrated with both high-grade calibration gases and a 3 L calibration syringe according to the recommendations by the manufacturer. Mask size was individually fitted prior to the test and maintained throughout the entire measurement.

Index device assessment

This study examined the validity of the Apple Watch Series 6 (AW), Garmin Fenix 6 (GF), and Huawei GT 2e (HW), all of

TABLE 1 Physical characteristics of participants.

	Male (n = 10)	Female (n = 10)	All participants (n = 20)
Age (yr)	23.8 ± 1.2	22.1 ± 2.7	23.0 ± 2.2
Height (cm)	178.5 ± 8.3	168.7 ± 8.5	173.6 ± 9.6
Weight (kg)	71.7 ± 10.3	62.6 ± 10.3	67.1 ± 11.1
BMI (kg/m ²)	22.4 ± 2.0	21.8 ± 1.9	22.1 ± 1.9
BF (%)	14.5 ± 4.8	23.8 ± 5.6	19.2 ± 6.9

Data are expressed as mean ± SD. BMI, body mass index; BF, body fat percentage; SD, standard deviation.

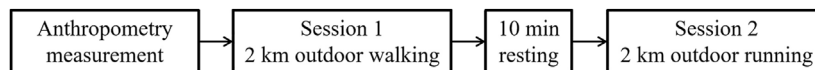


FIGURE 1

Experimental protocol. In the first session, the subjects were advised to walk a 2-km distance at approximately 6 km/h. In the second session, the subjects were asked to run a 2-km distance at approximately 10 km/h. Two sessions were separated by 10 minutes sitting break.

which were the newest generations introduced in January 2021. Each smartwatch has the outdoor walking and running mode, and records distance, speed and heart rate in real time. All devices use photoplethysmography to estimate heart rate from the wrist and use the GPS to estimate distance and speed during the outdoor walking and running. EE was estimated in real time and displayed.

All watches used in the study were bought commercially. Three watches were placed two on one wrist and one on the other wrist for different participants according to a pre-randomly allocation list in order to keep the number of devices on one wrist counterbalanced. The position of the watch was selected as careful as possible according to the manufacturer's instructions.

Protocol

Data for this study were collected during one visit. Subjects were instructed not to consume food, coffee, tea, or other stimulants e.g. energy drinks or soft drinks except water at least 6 h prior to measurements and not to do any vigorous physical activity and consume alcohol during 24 h prior to measurements. Subjects' height was measured to the nearest 0.1 cm using a wall-mounted height scale. Body mass and body composition were measured after emptying the bladder and in light underwear using a calibrated InBody 720 bio-impedance device (Biospace, Co, Ltd., Seoul, Korea). Height, weight, gender, and date of birth were used to initialize the smartwatches for each subject.

After the anthropometry measurement, the protocol included two exercise sessions performed on the outdoor

running track as showed in Figure 1. Two sessions were separated by a recovery period, which was 10 minutes sitting break. In the first session, the subjects were advised to walk a 2-km distance at approximately 6 km/h. In the second session, the subjects were asked to run a 2-km distance at approximately 10 km/h. Subjects were instructed to keep a steady pace during the session. The walking speed was 6.15 ± 0.29 km/h and the running speed was 10.94 ± 0.99 km/h. The examiner confirmed that heart rate had returned to resting levels before the second session. The corresponding outdoor walking or running mode was selected and started and ended at the same time for all smartwatches. The starting and end time of each session was recorded on the data sheet by researchers. The temperature and relative humidity was $16.5 \pm 2.6^\circ\text{C}$.and $56.7 \pm 14.3\%$ for walking and $13.9 \pm 2.9^\circ\text{C}$.and $62.6 \pm 21.8\%$ for running.

Data acquisition and processing

The criterion data from the K5 were downloaded breath-by-breath and included measures of EE. EE values from the K5 were summed individually for each exercise session. The EE estimates from the watches were obtained directly from screenshots of the respective applications, because the watches primarily reported cumulative totals. Due to technical errors, we only obtained 19 EE data from HW for walking and running, respectively.

Data analyses

Statistical analyses were conducted using IBM SPSS Statistics software version 24 (IBM, Armonk, New York).

TABLE 2 Descriptive examination of the differences between the estimated EE (smartwatches) and the measured EE (K5) during the outdoor walking and running.

Activity	Device	N	EE (kcal)	Diff (kcal)	MAPE (%)	ICC	t	p	p
Walking	AW	20	129.1 ± 20.1	20.5 ± 11.3	19.8 ± 12.4	0.821	-8.129	0.000	1.000
	GF	20	139.6 ± 39.6	31.0 ± 38.3	32.0 ± 34.1	0.216	-3.615	0.002	0.929
	HW	19	111.2 ± 18.1	2.0 ± 12.4	9.9 ± 8.2	0.760	-0.703	0.491	0.105
	K5	20	108.7 ± 17.4	NA	NA	NA	NA	NA	NA
Running	AW	20	137.8 ± 23.1	25.7 ± 16.0	24.4 ± 16.1	0.741	-7.162	0.000	0.999
	GF	20	137.8 ± 23.1	19 ± 22.8	21.8 ± 17.3	0.594	-3.715	0.001	0.942
	HW	19	111.8 ± 21.1	-0.3 ± 16.7	11.9 ± 9.9	0.698	0.069	0.946	0.051
	K5	20	112.2 ± 21.3	NA	NA	NA	NA	NA	NA

Data are expressed as mean ± SD, for EE, Diff and MAPE. Data from the smartwatches were compared with the criterion using paired sample t-tests. EE, energy expenditure; Diff: the difference of the estimated EE (smartwatches) with the measured EE (K5); K5, the golden standard of EE, assessment; AW, Apple Watch Series 6; GF, Garmin FENIX, 6; HW, Huawei Watch GT, 2e; MAPE, mean absolute percentage error; ICC, intraclass correlation coefficient; SD, standard deviation; p, statistical power.

Descriptive statistics were used to characterize the sample population. The validity of the watches was determined by several statistical tests. Data from the smartwatches were compared with the criterion using paired sample t-tests. The mean absolute percentage errors (MAPE) were calculated as an indicator of measurement error. The absolute percentage errors (%) were calculated as follows: $|EE \text{ from smartwatches} - EE \text{ from K5}| / (EE \text{ from K5}) \times 100$ for each subject. A MAPE of $\leq 10\%$ was used as the criterion value for validity (Nelson et al., 2016). As the commonly used method to validate wearable devices, the Intraclass Correlation Coefficient (ICC) defined the agreement between the gold standard and the tested devices, providing an estimate of overall concordance between two methods (Fokkema et al., 2017; Wahl et al., 2017; Boudreaux et al., 2018). Excellent, good, moderate, and low agreement thresholds were defined as ICC values of ont, goo). Excellent, goooices, providin (Fokkema et al., 2017). To investigate the level of agreement, Bland-Altman plots were prepared according to Bland and Altman (Bland and Altman, 1986). For this, limits of agreement were set to 95%.

Results

EE during the outdoor walking

The EE during the outdoor walking was showed in Table 2. On average, participants achieved an EE of 108.7 ± 17.4 kcal in K5 during the outdoor walking. The average estimated EE was 129.1 ± 20.1 kcal, 139.6 ± 39.6 kcal and 111.2 ± 18.1 kcal from AW, GF and HW, respectively. Paired t-test analysis showed that EE from AW and GF were significantly higher than that obtained by the K5 ($p < 0.001$ and 0.002 , respectively), but not significantly different for HW ($p = 0.491$). Compared to the K5, the MAPE of

EE was 19.8%, 32.0% and 9.9% in AW, GF and HW, respectively (Table 2). Individual EE values from AW (ICC = 0.821) and HW (ICC = 0.760) also showed good agreement with K5. However, the agreement for GF (ICC = 0.216) was low (Fokkema et al., 2017).

In the Bland-Altman plots, the percentage of the values within the limits of agreement was 100%, 90% and 95% for AW (Figure 2A), GF (Figure 2C) and HW (Figure 2E), respectively. The AW showed the narrowest limits of agreement (44.1 kcal), followed by HW of 48.6 kcal, while the GF exhibited the broadest limits of agreement (150.1 kcal).

EE during the outdoor running

The EE during the outdoor running was showed in Table 2. On average, participants achieved an EE of 112.2 ± 21.3 kcal in K5 during the outdoor running. The average estimated EE was 137.8 ± 23.1 kcal, 131.1 ± 28.8 kcal and 111.8 ± 21.1 kcal from AW, GF and HW, respectively. Paired t-test analysis showed that EE from AW and GF were significantly higher than the K5 ($p < 0.001$ and 0.001 , respectively), but not significantly different for HW ($p = 0.946$). Compared to the K5, the MAPE of EE was 24.4%, 21.8% and 11.9% in AW, GF and HW, respectively. Individual EE values from AW (ICC = 0.741) and HW (ICC = 0.698) also showed moderate agreement with K5. However, the agreement for GF (ICC = 0.594) was low (Fokkema et al., 2017).

In the Bland-Altman plots, the percentage of the values within the limits of agreement was 95% for AW (Figure 2B) and GF (Figure 2D) and 100% for HW (Figure 2F). The AW showed the narrowest limits of agreement (62.8 kcal), followed by HW of 65.6 kcal, whereas the GF exhibited the broadest limits of agreement (89.4 kcal).

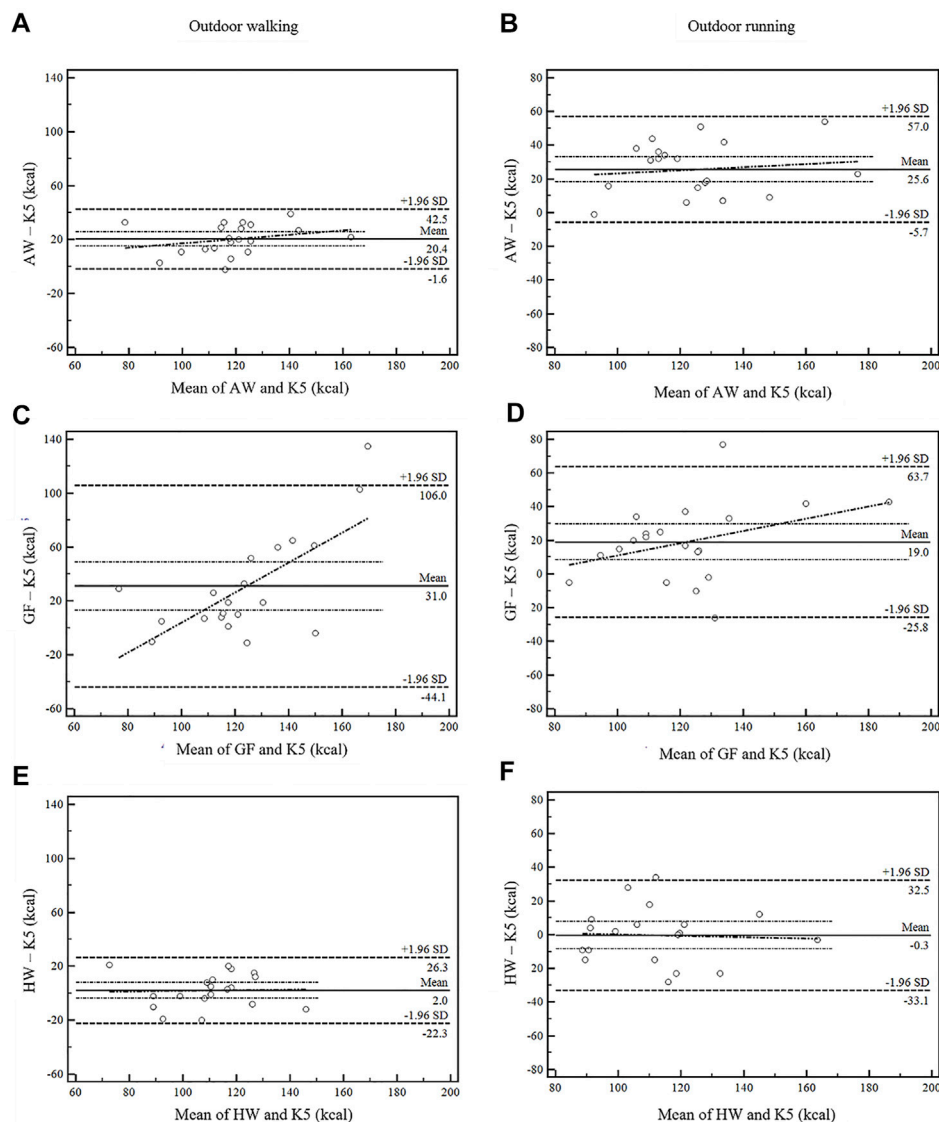


FIGURE 2

Bland-Altman plots comparing the EE estimations by smartwatches (AW (A,B), GF (C,D), HW (E,F)) and K5. The differences of the EE values on the y-axis relative to the mean of the two methods (smartwatch and K5) on the x-axis. Mean differences (bias) between estimated EE and EE of criterion, upper and lower limits of agreement (ULoA, LLoA) are labeled in the plots. Limits of agreement (LoA) were calculated as means ± 1.96 SD. K5, the criterion of EE assessment; AW, Apple Watch Series 6; GF, Garmin FENIX 6; HW, Huawei Watch GT 2e; SD, standard deviation.

Discussion

This study investigated the validity of three popular smartwatches (AW, GF, HW) for estimating EE during outdoor walking and running by means of a portable gas analysis system (K5) as the criterion. The results revealed that some, but not all, of the smart-watches provide a reasonably accuracy for EE estimation. Estimations of EE provided by AW and HW demonstrated good/moderate criterion agreement, while the EE from GF showed poor agreement.

Although the HW yielded the best overall results, the AW also performed well in this study. The HW showed the lowest MAPE (9.9% for walking and 11.9% for running), while AW had the narrowest limits of agreement (44.1 kcal for walking and 62.8 kcal for running). Indeed, a MAPE of <10% can be considered as a high accuracy (Nelson et al., 2016). Since the manufacturer of HW is based in China and a Chinese population was included in this study it is likely that this explains the high accuracy. In a previous study it was shown that the accuracy of EE estimations by activity monitors differed between different ethnicities (Brazeau et al., 2014), but in that study only

accelerometer data was used for EE estimations. Unfortunately, most manufacturers do not provide any details on the algorithms that are used for EE estimations and as such, it remains unknown which variables were used in the devices tested in the present study. However, it is likely that a combination of accelerometry, PPG and GPS derived data was utilized. As such, we are only able to assess the ecological validity that is the result of both sensor quality and algorithm but the actual source of inaccuracy especially observed in GF remains unknown.

There is considerable variability of device accuracy for predicting EE of different types of activities (O’driscoll et al., 2020; Lee et al., 2014). The present study demonstrated that overall error estimates were similar between walking and running for AW and HW, but different for GF. GF could likely provide better EE estimates for the outdoor walking (MAPE = 32.0%) than running (MAPE = 21.8%). Wahl et al. (Wahl et al., 2017) examined the influence of running pace on the EE. The study shows a significant influence of running pace on the estimated EE. Energy expenditure tends to be overestimated at lower pace and underestimated at higher pace. Roos et al. (Roos et al., 2017) found that sportswatches significantly underestimated EE during the high intensity running with a proportional error increasing as the running speed increased. Both studies suggested improving the EE estimation algorithm in all range speed (Roos et al., 2017; Wahl et al., 2017). Nevertheless, this diversity also may be related to inaccuracies of the sensor (e.g. movement artefacts) at different speed. These observations indicated that consumer/health professionals should take more attention to select the most accurate device for the distinct types of activity.

Smartwatch EE overestimated EE versus the K5 during the outdoor walking and running. These observations are congruent with previous studies (Bai et al., 2016; Pope et al., 2019). EE during outdoor walking was overestimated by 19.8 % and 32.0% by the AW and GF, respectively, while the energy cost of running was overestimated by 24.4 % and 21.8%. Considering smartwatch EE data may be used for the purpose of maintaining or even losing body weight, these observations are remarkable. For example, the individual may set the goal of increasing the EE by 250 kcal per day through walking or running in order to lose 1 kg body weight per month. However, due to the smartwatch’s EE data inaccuracy, the actual increasing daily EE may be less than 200 kcal, resulting in the ineffectiveness of the weight loss program. Our findings, therefore, suggest a cautious approach to be taken when interpreting smartwatch EE data given the observed EE overestimation.

When interpreting our findings, one needs to be bear in mind a few limitations. Most notably, our study was designed to include only outdoor walking and running with advised self-selected steady speed. Furthermore, our findings may not translate directly into other activities due to e.g. differences in motion artefacts. In addition, our sample was restricted to healthy young Chinese within the normal range of BMI.

Therefore, we cannot generalize these findings to the other diverse population groups, such as elderly or obese. Furthermore, as the body side was randomized, we side specific effects of the accuracy. Future studies should consider powering the study to compare the effects of different wearing locations on the accuracy of the devices. Finally, this study did not assess the reliability of EE estimation from the smartwatches. Despite issues with accuracy, the reliability is very important for individuals and health professionals to monitor changes in EE over time. Future studies should incorporate reliability testing as well.

Conclusion

The findings of this study indicate that the smartwatches of AW, GF and HW may generally have moderate validity in EE estimates for outdoor walking and running. Small sample size with large variables even had sufficient power still a limit factor to generalize the results for different populations. Consumers should also be cautious when using the tested smart-watches for prediction of their energy expenditure.

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 the Institutional Review Board for Human Research Protections of Shanghai Jiao Tong University (registration number B2020024I, 2020.09.03). The patients/participants provided their written informed consent to participate in this study.

Author contributions

SL and WY designed and supervised the overall study. XW, TZ and SML collected the data. SL, SC and MS conducted the data analysis. SL and MS prepared the manuscript. All authors reviewed and approved the final manuscript.

Funding

This research was supported by Huawei Terminal Co. LTD. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the

decision to submit it for publication. All authors declare no other competing interests.

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.

References

- Argent, R., Hetherington-Rauth, M., Stang, J., Tarp, J., Ortega, F. B., Molina-Garcia, P., et al. (2022). Recommendations for determining the validity of consumer wearables and smartphones for the estimation of energy expenditure: Expert statement and checklist of the INTERLIVE network. *Sports Med.* 52, 1817–1832. doi:10.1007/s40279-022-01665-4
- Bai, Y., Welk, G. J., Nam, Y. H., Lee, J. A., Lee, J. M., Kim, Y., et al. (2016). Comparison of consumer and research monitors under semistructured settings. *Med. Sci. Sports Exerc.* 48, 151–158. doi:10.1249/MSS.0000000000000727
- Bland, J. M., and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310. doi:10.1016/s0140-6736(86)90837-8
- Boudreaux, B. D., Hebert, E. P., Hollander, D. B., Williams, B. M., Cormier, C. L., Naquin, M. R., et al. (2018). Validity of wearable activity monitors during cycling and resistance exercise. *Med. Sci. Sports Exerc.* 50, 624–633. doi:10.1249/MSS.0000000000001471
- Brazeau, A. S., Suppere, C., Strychar, I., Belisle, V., Demers, S. P., and Rabasa-Lhoret, R. (2014). Accuracy of energy expenditure estimation by activity monitors differs with ethnicity. *Int. J. Sports Med.* 35, 847–850. doi:10.1055/s-0034-1371837
- Charlot, K., Cornolo, J., Borne, R., BrugniauxandPichon, J. V. A., Richalet, J. P., Chapelot, D., et al. (2014). Improvement of energy expenditure prediction from heart rate during running. *Physiol. Meas.* 35, 253–266. doi:10.1088/0967-3334/35/2/253
- Deblois, J. P., White, L. E., and Barreira, T. V. (2021). Reliability and validity of the COSMED K5 portable metabolic system during walking. *Eur. J. Appl. Physiol.* 121, 209–217. doi:10.1007/s00421-020-04514-2
- Evenson, K. R., Goto, M. M., and Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int. J. Behav. Nutr. Phys. Act.* 12, 159. doi:10.1186/s12966-015-0314-1
- Fokkema, T., Kooiman, T. J., Krijnen, W. P., Cees, V. D. S., Martijn, D. E. G., and De Groot, M. (2017). Reliability and validity of ten consumer activity trackers depend on walking speed. *Med. Sci. Sports Exerc.* 49, 793–800. doi:10.1249/MSS.0000000000001146
- Jakicic, J. M., Marcus, M., Gallagher, K. I., Randall, C., Thomas, E., Goss, F. L., et al. (2004). Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Med. Sci. Sports Exerc.* 36, 897–904. doi:10.1249/01.mss.0000126805.32659.43
- Jonas, S. (2009). *ACSM's exercise is medicine : a clinician's guide to exercise prescription*. Philadelphia: Lippincott, Williams and Wilkins.
- Kinnunen, H., Hakkinen, K., Schumann, M., Karavirta, L., Westerterp, K. R., and Kyrolainen, H. (2019). Training-induced changes in daily energy expenditure: Methodological evaluation using wrist-worn accelerometer, heart rate monitor,

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.

and doubly labeled water technique. *PLoS ONE* 14, e0219563. doi:10.1371/journal.pone.0219563

Laricchia, F. (2022). Total wearable device unit shipments worldwide 2014-2021 Available at: <https://www.statista.com/statistics/437871/wearables-worldwide-shipments/> [Accessed March 29, 2022].

Lee, J. M., Kim, Y., and Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Med. Sci. Sports Exerc.* 46, 1840–1848. doi:10.1249/MSS.0000000000000287

Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., and Rowland, J. L. (2014). Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *J. Med. Internet Res.* 16, e192. doi:10.2196/jmir.3469

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., and Montoye, A. H. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Med. Sci. Sports Exerc.* 48, 1619–1628. doi:10.1249/MSS.0000000000000933

O'driscoll, R., Turicchi, J., Beaulieu, K., Scott, S., Matu, J., Deighton, K., et al. (2020). How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *Br. J. Sports Med.* 54, 332–340. doi:10.1136/bjsports-2018-099643

Passler, S., Bohrer, J., Blochinger, L., and Senner, V. (2019). Validity of wrist-worn activity trackers for estimating VO2max and energy expenditure. *Int. J. Environ. Res. Public Health* 16.E3037. doi:10.3390/ijerph16173037

Pope, Z. C., Lee, J. E., Zeng, N., and Gao, Z. (2019). Validation of four smartwatches in energy expenditure and heart rate assessment during exergaming. *Games Health J.* 8, 205–212. doi:10.1089/g4h.2018.0087

Roos, L., Taube, W., Beeler, N., and Wyss, T. (2017). Validity of sports watches when estimating energy expenditure during running. *BMC Sports Sci. Med. Rehabil.* 9, 22. doi:10.1186/s13102-017-0089-6

Schneller, M. B., Pedersen, M. T., Gupta, N., Aadahl, M., and Holtermann, A. (2015). Validation of five minimally obstructive methods to estimate physical activity energy expenditure in young adults in semi-standardized settings. *Sensors (Basel)* 15, 6133–6151. doi:10.3390/s150306133

Thompson, W. R. (2021). Worldwide survey of fitness trends for 2021. *ACSM's Health Fit. J.* 25, 10–19. doi:10.1249/Fit.0000000000000631

Van Hove, K., Boen, F., and Lefevre, J. (2015). Validation of the SenseWear Armband in different ambient temperatures. *J. Sports Sci.* 33, 1007–1018. doi:10.1080/02640414.2014.981846

Wahl, Y., Dukung, P., Droszez, A., Wahl, P., and Mester, J. (2017). Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. *Front. Physiol.* 8, 725. doi:10.3389/fphys.2017.00725