

University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Health & Human Performance Faculty
Publications and Presentations

College of Health Professions

4-2018

Validity of wrist-worn consumer products to measure heart rate and energy expenditure

R. S. Thiebaud

Merrill D. Funk

Jacelyn C. Patton

Brook L. Massey

Terri E. Shay

See next page for additional authors

Follow this and additional works at: https://scholarworks.utrgv.edu/hhp_fac



Part of the [Kinesiology Commons](#), and the [Public Health Commons](#)

Authors

R. S. Thiebaud, Merrill D. Funk, Jacelyn C. Patton, Brook L. Massey, Terri E. Shay, Martin G. Schmidt, and Nicolas Giovannitti

Validity of wrist-worn consumer products to measure heart rate and energy expenditure

Digital Health
Volume 4: 1–7
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2055207618770322
journals.sagepub.com/home/dhj



Robert S. Thiebaud¹, Merrill D. Funk², Jacelyn C. Patton¹, Brook L. Massey¹,
Terri E. Shay¹, Martin G. Schmidt¹ and Nicolas Giovannitti¹

Abstract

Introduction: The ability to monitor physical activity throughout the day and during various activities continues to improve with the development of wrist-worn monitors. However, the accuracy of wrist-worn monitors to measure both heart rate and energy expenditure during physical activity is still unclear. The purpose of this study was to determine the accuracy of several popular wrist-worn monitors at measuring heart rate and energy expenditure.

Methods: Participants wore the TomTom Cardio, Microsoft Band and Fitbit Surge on randomly assigned locations on each wrist. The maximum number of monitors per wrist was two. The criteria used for heart rate and energy expenditure were a three-lead electrocardiogram and indirect calorimetry using a metabolic cart. Participants exercised on a treadmill at 3.2, 4.8, 6.4, 8 and 9.7 km/h for 3 minutes at each speed, with no rest between speeds. Heart rate and energy expenditure were manually recorded every minute throughout the protocol.

Results: Mean absolute percentage error for heart rate varied from 2.17 to 8.06% for the Fitbit Surge, from 1.01 to 7.49% for the TomTom Cardio and from 1.31 to 7.37% for the Microsoft Band. The mean absolute percentage error for energy expenditure varied from 25.4 to 61.8% for the Fitbit Surge, from 0.4 to 26.6% for the TomTom Cardio and from 1.8 to 9.4% for the Microsoft Band.

Conclusion: Data from these devices may be useful in obtaining an estimate of heart rate for everyday activities and general exercise, but energy expenditure from these devices may be significantly over- or underestimated.

Keywords

Photoplethysmography, physical activity, fitness trackers, activity monitors, Fitbit

Submission date: 30 August 2017; Acceptance date: 8 March 2018

Introduction

The development of wearable technology to track both heart rate and energy expenditure has improved over the past few years due to the use of photoplethysmography. Data from these devices may facilitate healthy behaviors such as increased physical activity.^{1,2} However, if the information collected from wearable technology is not accurate, the usefulness of these devices is limited. Photoplethysmography monitors heart rate by using light emitting diodes and a photo diode.³ Shorter wavelengths (green light) are emitted into the skin to help minimize motion artifacts, but it does not penetrate skin depth as well as longer wavelengths.³ Despite using shorter wavelength light, motion artifacts are

still present with this technology, so different algorithms are used to further decrease motion artifacts.³ If heart rates are monitored accurately, they could be used to track exercise intensity and improve the estimation of energy expenditure. For proprietary reasons, many technology companies do not reveal how they validate their technology or which variables they use to estimate energy expenditure, so it is difficult for consumers to

¹Department of Kinesiology, Texas Wesleyan University, USA

²Department of Health and Human Performance, University of Texas Rio Grande Valley, USA

Corresponding author:

Robert S. Thiebaud, 1201 Wesleyan Street, Fort Worth, TX 76105, USA.
Email: rthiebaud@txwes.edu



know how valid different devices are and how they compare with other devices.

Few studies have investigated the validity of wrist-worn devices that use photoplethysmography to measure both heart rate and energy expenditure.^{4–7} Some studies have exclusively examined heart rate and found sufficient accuracy during treadmill walking and running,^{8,9} while others found less accurate readings.^{10,11} Many different devices and protocols have been used in previous studies, therefore it is important to build sufficient evidence to provide consumers with valuable information on the validity of popular devices. The purpose of this study was to determine the validity of three common wrist-worn consumer monitors at measuring heart rate and energy expenditure during walking, jogging and running. We hypothesized that the devices would be more accurate in estimating heart rate compared to energy expenditure.

Methods

Participants

Twenty recreationally active males and two females participated in this study (mean (SD): age = 22 (3) years, height = 1.73 (0.09) m and weight = 75.9 (10.2) kg). Participants were told about the nature, purpose, details and any risks associated with the experiment, and each participant gave written informed consent. The University's Institutional Review Board approved the protocol of the research study.

Exercise protocol

Wrist-worn monitors were randomly placed on subjects' wrists with a maximum of two monitors on one wrist. Devices were randomly placed on the wrists to avoid any bias that may be produced by placing devices in the same place each time. A possibility of less accurate readings may occur with more than one monitor on a wrist, but monitors were placed on the wrist based on the manufacturer's instructions. Other studies have also used similar procedures to test the validity of these devices.^{6–8} Participants exercised on a treadmill at 3.2, 4.8, 6.4, 8 and 9.7 km/h for 3 minutes at each speed with no rest between speeds. These speeds were chosen to reflect various intensities that the general healthy population may experience, and these speeds have been used in other studies.^{8,11} The duration of 3 minutes was chosen to allow heart rate to reach steady state at each intensity. Other studies have also used 3–5-minute stages.^{7,8,11} Heart rate was measured using electrocardiography (three-lead electrocardiogram (ECG), Quinton[®] Q-Stress, version 4.5, Cardiac Science, Bothell, WA, USA). Energy expenditure was

measured using a metabolic cart system (Trueone 2400[®] metabolic measurement system, Parvomedics, Sandy, UT, USA).

Consumer wrist-worn monitors

Fitbit Surges (Fitbit Inc., San Francisco, CA, USA), Microsoft Bands (Microsoft Corp., Redmond, WA, USA), and TomTom Cardios (TomTom Inc., Burlington, MA, USA) were placed on the wrist and set to "treadmill" mode if available and according to the manufacturers' recommendations.

Statistical analysis

The average heart rate and energy expenditure recorded during the 3 minutes at each speed were used for analysis. Pearson correlations measured associations between criterion variables and wrist-worn monitors. Spearman's rank correlation coefficients were used for any variables that were not normally distributed. Statistical significance was set at an α level of <0.01 . The criterion measure for heart rate was the ECG and for energy expenditure was the metabolic cart. Mean bias was calculated by subtracting the wrist-worn device from the criterion and 95% limits of agreement were also calculated. For equivalence testing, 95% precision was assumed if the wrist-worn monitors' 90% confidence intervals were within an equivalence zone that was between $\pm 10\%$ of the criterion mean for energy expenditure and $\pm 5\%$ of the criterion mean for heart rate.¹² Mean absolute percentage error (MAPE) $((\text{monitor} - \text{criterion})/\text{criterion} \times 100\%)$ provided a general measurement error for monitors.

Results

Heart rate

Wrist-worn monitors overestimated heart rates compared to the criterion for all speeds except for the Fitbit Surge, which underestimated heart rate at 8 and 9.7 km/h (Table 1). The MAPE varied from 2.17 to 8.06% for the Fitbit Surge, from 1.01 to 7.49% for the TomTom Cardio and from 1.31 to 7.37% for the Microsoft Band (Table 1). The equivalence zones for heart rate are found in figure 1.

Energy expenditure

The Fitbit Surge overestimated energy expenditure at each speed, while the TomTom Cardio overestimated energy expenditure at 3.2, 4.8 and 6.4 km/h and underestimated energy expenditure at 8 km/h and 9.7 km/h. The Microsoft Band underestimated energy expenditure

Table 1. Heart rate.

	Heart rate (bpm)			
	Fitbit Surge	TomTom Cardio	Microsoft Band	ECG
3.2 km/h				
Mean (SD)	89 (11)	85 (10)	89 (10)	84 (10)
Mean bias (SD)	5 (19)	1 (9)	6 (8)	
LoA	70–108	76–94	73–106	
Correlation	0.57 ^a	0.89 ^a	0.62 ^a	
MAPE (%)	6.5 (13.13)	1.01 (1.38)	7.37 (10.75)	
Equivalence zones	85–91	81–87 ^b	85–91	76–92 ^c , 80–88 ^d
4.8 km/h				
Mean (SD)	100 (13)	98 (14)	97 (8)	91 (9)
Mean bias (SD)	7 (10)	7 (12)	5 (10)	
LoA	80–120	75–120	77–116	
Correlation	0.53	0.55 ^a	0.35	
MAPE (%)	8.06 (12.04)	7.49 (13.41)	6.59 (11.86)	
Equivalence zones	95–103	92–100	94–99	82–100 ^c , 87–96 ^d
6.4 km/h				
Mean (SD)	114 (11)	116 (13)	114 (10)	112 (9)
Mean bias (SD)	2 (8)	5 (9)	3 (8)	
LoA	98–130	100–133	98–130	
Correlation	0.69 ^a	0.76 ^a	0.63 ^a	
MAPE (%)	2.17 (7.68)	4.5 (7.45)	2.46 (7.07)	
Equivalence zones	110–116 ^b	111–119	109–115 ^b	100–122 ^c , 106–117 ^d
8 km/h				
Mean (SD)	132 (13)	141 (12)	141 (11)	135 (10)
Mean bias (SD)	−4 (8)	6 (11)	5 (12)	
LoA	117–147	120–162	118–164	
Correlation	0.82 ^a	0.52	0.37	
MAPE (%)	−2.77 (6.04)	4.38 (8.42)	4.19 (9.15)	
Equivalence zones	127–135	137–144	137–143 ^b	122–150 ^c , 129–143 ^d
9.7 km/h				
Mean (SD)	150 (15)	157 (13)	156 (13)	155 (13)

(continued)

Table 1. Continued.

	Heart rate (bpm)			
	Fitbit Surge	TomTom Cardio	Microsoft Band	ECG
Mean bias (SD)	−5 (9)	2 (6)	2 (13)	
LoA	133–167	146–167	132–181	
Correlation	0.84 ^a	0.91 ^a	0.53	
MAPE (%)	−3.35 (5.51)	1.11 (5.52)	1.31 (8.05)	
Equivalence zones	145–153	153–160 ^b	153–160 ^b	140–171 ^c , 147–163 ^d

Values are mean (SD).

ECG: electrocardiogram; LoA: 95% limits of agreement; MAPE: mean absolute percentage error.

^a*p*-value < 0.01.

^bIndicates that values are within the 5% equivalence zone of the electrocardiogram.

^cIndicates 10% equivalence area.

^dIndicates 5% equivalence area.

Table 2. Energy expenditure.

	Energy expenditure (kcal)			
	Fitbit Surge	TomTom Cardio	Microsoft Band	MetCart
3.2 km/h				
Mean (SD)	8.7 (2.3)	6.2 (2.2)	5.4 (1.6)	6.1 (1.0)
Mean bias (SD)	2.7 (2.0)	0.1 (1.8)	−0.7 (1.7)	
LoA	4.9–12.5	2.7–9.6	2–8.8	
Correlation	0.52	0.57 ^a	0.19	
MAPE (%)	44.5 (33.0)	0.4 (31.6)	−9.4 (27.9)	
Equivalence zone	7.9–9.2	5.4–6.7	4.8–5.7	5.5–6.7
4.8 km/h				
Mean (SD)	27.8 (5.4)	19.3 (5.2)	16.2 (4.0)	17.1 (2.9)
Mean bias (SD)	10.3 (4.4)	2.2 (3.7)	−0.9 (4.3)	
LoA	19.2–36.4	11.9–26.6	7.8–24.7	
Correlation	0.53	0.70 ^a	0.24	
MAPE (%)	61.8 (27.5)	12.0 (22.2)	−3.6 (24.1)	
Equivalence zones	25.9–29.0	17.5–20.4	14.8–17.1	15.4–18.8
6.4 km/h				
Mean (SD)	51.0 (8.1)	34.4 (9.1)	32.1 (6.2)	33.4 (5.7)
Mean bias (SD)	17 (6.3)	1.0 (6.6)	−1.3 (7.7)	

(continued)

Table 2. Continued.

	Energy expenditure (kcal)			
	Fitbit Surge	TomTom Cardio	Microsoft Band	MetCart
LoA	38.6–63.4	17.8–47.4	47.2	
Correlation	0.55 ^a	0.66 ^a	0.15	
MAPE (%)	52.7 (21.8)	2.5 (20.3)	–1.8 (22.3)	
Equivalence zones	48.2–52.7	31.1–36.4 ^b	29.9–33.4	30.1–36.7
8 km/h				
Mean (SD)	80 (11.6)	49.5 (12.7)	59.7 (10.5)	57.9 (10.0)
Mean bias (SD)	20.7 (9.8)	–9.5 (9.8)	1.1 (12.8)	
LoA	60.8–99.2	29.7–68.4	34.6–84.7	
Correlation	0.53	0.60 ^a	0.20	
MAPE (%)	37.0 (19.1)	–16.4 (16.8)	4.0 (21.6)	
Equivalence zones	75.8–82.5	45.0–52.2	56.3–62.3 ^b	52.1–63.7
9.7 km/h				
Mean (SD)	112.7 (16.1)	66.7 (17.0)	96.1 (16.4)	90.7 (15.1)
Mean bias (SD)	21.7 (13.3)	–24.0 (13.4)	5.4 (17.4)	
LoA	86.6–138.7	40.5–92.9	61.9–130.3	
Correlation	0.60 ^a	0.56 ^a	0.37	
MAPE (%)	25.4 (15.7)	–26.6 (15.0)	7.5 (18.6)	
Equivalence zones	106.9–116.2	60.7–70.3	90.3–99.6 ^b	81.6–99.8

Values are mean (SD).

MetCart: metabolic cart; LoA: 95% limits of agreement; MAPE: mean absolute percentage error.

^a*p*-value < 0.01.

^bIndicates values are within the equivalence zone of the metabolic cart.

at 3.2, 4.8 and 6.4 km/h and overestimated energy expenditure at 8 and 9.7 km/h (Table 2). The MAPE varied from 25.4 to 61.8% for the Fitbit Surge, from 0.4 to 26.6% for the TomTom Cardio and from 1.8 to 9.4% for the Microsoft Band (Table 2). The equivalence zones for energy expenditure are found in figure 2.

Discussion

The main findings from this study were that wrist-worn monitors produce more accurate readings for heart rates compared to energy expenditure. However, the accuracy of the devices may be influenced by the intensity.

When comparing the accuracy of heart rates from the wrist-worn monitors to the ECG readings,

the highest correlations were at the fastest speeds for the Fitbit Surge ($r=0.84$) and TomTom Cardio ($r=0.91$), while the highest correlation for the Microsoft Band was at 6.4 km/h ($r=0.63$). Stahl et al.⁸ performed a similar study to ours and found higher correlations than we did for the TomTom Cardio ($r=0.959$) and Microsoft Band ($r=0.956$), although they used the average heart rates throughout the entire exercise protocol to determine their correlations. In another study, Gillinov et al.¹⁰ found concordance correlations of 0.88 between the TomTom Surge and ECG leads. Part of the reason for the lower correlations between heart rates in the current study may be due to a smaller sample size, a different criterion measure used (ECG vs. polar heart

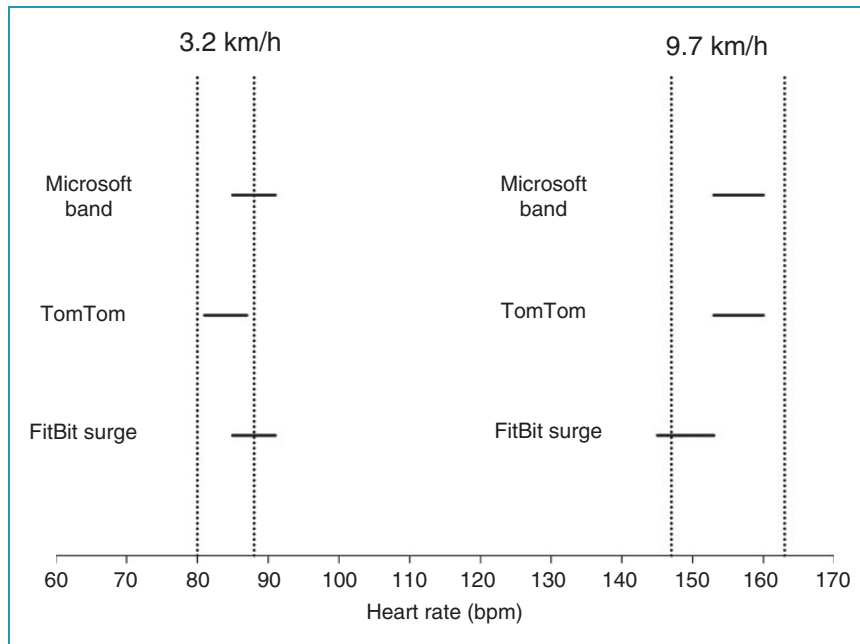


Figure 1. Heart rate equivalence testing to evaluate agreement between devices and electrocardiogram (ECG) at 3.2 and 9.7 km/h. Dashed lines represent the 5% equivalence zones for ECG and solid lines represent 90% confidence intervals for different devices.

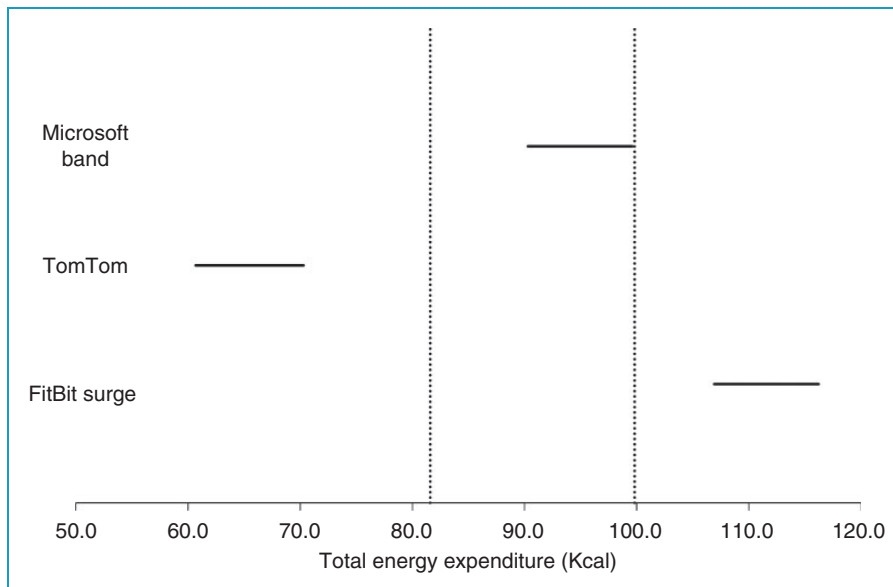


Figure 2. Equivalence testing for total energy expenditure at 9.7 km/h to evaluate agreement between devices and metabolic cart. Dashed lines represent the 10% equivalence zones for the metabolic cart and solid lines represent 90% confidence intervals for different devices.

rate monitor) and that we correlated heart rates at each speed instead of an overall heart rate.

When examining the MAPE, other studies have found similar results to the current study. For example, Stahl et al.⁸ found that MAPE varied from 0.97 to 5.71% for the TomTom Cardio and from 3.06 to 8.39% for the Microsoft Band, while Gillinov et al.¹⁰ found MAPE of 6.2% for the TomTom Cardio.

Similarly, we found that MAPE varied from 1.01 to 7.49% for the TomTom Cardio and from 1.31 to 7.37% for the Microsoft Band, with the lower MAPE found at the faster speeds. Shcherbina et al.⁶ also found a larger percent error during walking compared to running for the Fitbit Surge and Microsoft Band. Overall, it appears that the faster the heart rate due to increasing speed, the more accurate the devices become.

Although heart rate was fairly accurate using these monitors, energy expenditure varied much more and did not necessarily correlate with heart rate. For example, when examining the total energy expenditure, MAPE was greater than 20% for the Fitbit Surge and TomTom Cardio, while that of the Microsoft Band was only 7.5%. This confirms other studies that used similar wrist-worn monitors in that they do not accurately measure energy expenditure.^{5–7} Though it is unclear how each device calculates the energy expenditure, it does not appear that monitoring heart rate concurrently creates an accurate measure of energy expenditure.

One factor that could have impacted the results was the use of a treadmill mode. Two of the devices had a treadmill mode while one device did not. This may have limited some of the ability to accurately measure heart rate and energy expenditure in the device without a treadmill mode. The impact of this on the results is unclear and future studies should determine how using different modes influences the measurement of heart rate and energy expenditure. In addition, because participants only exercised through the multiple speeds once, the reliability of the measurements is unclear.

One limitation of the study was the small sample size. A smaller sample size can lead to a lack of uniformity and can decrease statistical power. Despite the small sample size and large variation in the current study, it appears that the results follow a pattern similar to those of other studies investigating these devices.^{8,10}

Conclusions

Wrist-worn monitors report more accurate heart rates than energy expenditure during treadmill exercise. However, the accuracy of the devices for measuring heart rate may not yet be high enough for use in a research setting or for athletes who use heart rate measurement to reach precise heart rates for training purposes. Data from these devices may be useful in obtaining an estimate of heart rate for everyday activities and general exercise, but caution should be taken when using energy expenditure from these devices as the calories may be significantly over- or underestimated.

Acknowledgements: We would like to thank the participants who volunteered for this study.

Contributorship: RT and MF researched literature and conceived the study. RT, MF and JP were involved in protocol development and gaining ethical approval. RT, JP, BM, MS, TS and NG were involved in patient recruitment and data collection. RT and MF were involved in data analysis. RT wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Declaration of Conflicting Interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The ethics committee of Texas Wesleyan University approved this study (Research Ethics Committee number: SP170021).

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Sam Taylor Fellowship Award.

Guarantor: RT.

Peer review: This manuscript was reviewed by two individuals. The author(s) have elected for these individuals to remain anonymous.

References

1. Lewis ZH, Lyons EJ, Jarvis JM, et al. Using an electronic activity monitor system as an intervention modality: A systematic review. *BMC Public Health* 2015; 15: 585.
2. Patel MS, Asch DA and Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* 2015; 313: 459–460.
3. Zhou C, Feng J, Hu J, et al. Study of artifact-resistive technology based on a novel dual photoplethysmography method for wearable pulse rate monitors. *J Med Syst* 2016; 40: 56.
4. Parak J, Uuskoski M, Macheck J, et al. Estimating heart rate, energy expenditure, and physical performance with a wrist photoplethysmographic device during running. *JMIR Mhealth Uhealth* 2017; 5: e97.
5. Wallen MP, Gomersall SR, Keating SE, et al. Accuracy of heart rate watches: Implications for weight management. *PLoS One* 2016; 11: e0154420.
6. Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J Pers Med* 2017; 7: 3.
7. Dooley EE, Golaszewski NM and Bartholomew JB. Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR Mhealth Uhealth* 2017; 5: e34.
8. Stahl SE, An H, Dinkel DM, et al. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exerc Med* 2016; 2: e000106.
9. Delgado-Gonzalo R, Parak J, Tarniceriu A, et al. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. *Conf Proc IEEE Eng Med Biol Soc* 2015; 2015: 430–433.
10. Gillinov S, Etiwy M, Wang R, et al. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Med Sci Sports Exerc* 2017; 49: 1697–1703.
11. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol* 2017; 2: 104–106.
12. Chowdhury EA, Western MJ, Nightingale TE, et al. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PLoS one* 2017; 12: e0171720.