European Journal of Sport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tejs20

Running power: lab based *vs*. portable devices measurements and its relationship with aerobic power

Paolo Taboga, Nicola Giovanelli, Enrico Spinazzè, Francesco Cuzzolin, Giuseppe Fedele, Silvano Zanuso & Stefano Lazzer

EUROPEAN JOURNAL OF SPORT SCIENCE
https://doi.org/10.1080/17461391.2021.1966104

Running power: lab based *vs*. portable devices measurements and its relationship with aerobic power

Paolo Taboga [a], Nicola Giovanelli [b,c], Enrico Spinazzè[b,c], Francesco Cuzzolin[b], Giuseppe Fedele[d], Silvano Zanuso[d] and Stefano Lazzer [b,c]

[a]Department of Kinesiology, California State University, Sacramento, CA, USA; [b]Department of Medicine, University of Udine, Udine, Italy; [c]School of Sport Sciences, University of Udine, Udine, Italy; [d]Technogym Scientific Department, Technogym S.p.A, Cesena, Italy

ABSTRACT
In recent years, different companies have developed devices that estimate "running power". The main objective of this paper is to evaluate the effect of running speed on aerobic and running powers measured using force plates and by different devices. The second objective is to evaluate the relationship between aerobic power and running powers measured using force plates and by different devices. We enrolled 11 subjects in the study, they performed 5-min running trials at 2.22, 2.78, 3.33, 3.89 and 4.44 m/s respectively on a force-measuring treadmill while we collected metabolic data. We calculated running power as the dot product of ground reaction force and velocity of the centre of mass and compared it to the running power estimates of three devices: Skillrun (Technogym), Stryd Summit Powermeter (Stryd) and Garmin HRM-Run (Garmin). We found statistically significant linear correlations with running powers measured by all devices and running speed. Although absolute running power measurements were different among devices, an increase of 1 m/s in running speed translated to an increase of 0.944 W/kg in running power ($p < 0.001$). We found statistically significant linear correlations with running powers measured by all devices and aerobic power, in particular: as aerobic power increases by 1 W/kg, running power increases by 0.218 W/kg for all devices ($p < 0.001$). For level treadmill running, across speeds, running power measured by commercially available devices reflects force-based measurements and it can be a valuable metric, providing quasi real-time feedback during training sessions and competitions.

KEYWORDS
Energy cost of running; metabolic power; mechanical power; running power; running mechanics

Highlights

- We evaluated the effect of running speed on aerobic and running powers measured using force plates and by different devices.
- We also compared the relationship between aerobic power and running powers measured using force plates and by different devices.
- We found statistically significant linear correlations with running powers measured by all devices and aerobic power, in particular: as aerobic power increases by 1 W/kg, running power increases by 0.218 W/kg for all devices.
- For level treadmill running, across speeds, running power measured by commercially available devices reflects force-based measurements and it can be a valuable metric, providing quasi real-time feedback during training sessions and competitions.

Introduction

Mechanical power is a parameter commonly used in cycling to evaluate athletes (in terms of physiological characteristics, efficiency, etc.) and in training or racing settings (to design workouts, monitor effort, etc.) (Capostagno, Lambert, & Lamberts, 2016). Conversely, in running, athletes and coaches use pace (min/km) or heart rate to monitor training or racing intensity. Nevertheless, these parameters are affected by a number of factors (weather, fatigue, sleep quality …) making them difficult to interpret in some situations (Rodriguez, Brown, & Troped, 2005). Recently, to replicate the insight provided by mechanical power to cyclists (Paton & Hopkins, 2001), new devices have been developed for runners (Jaen-Carrillo, Roche-Seruendo, Cartón-Llorente, Ramírez-Campillo, & García-Pinillos, 2020). Different companies have put on the market wearable devices that estimate mechanical power during running or, as it is commonly referred to, "running power". These devices rely on inertial measurement units (IMUs) and global positioning system (GPS) sensors to measure the motions of the runner and, by means of proprietary

CONTACT Paolo Taboga paolo.taboga@csus.edu

(undisclosed) algorithms, provide a value of running power. The integration of IMUs + GPS sensors has been reported to overcome the intrinsic limits of GPS sensors alone when trying to accurately determine position and velocity (Townshend, Worringham, & Stewart, 2008) and estimate energy expenditure (Hongu, Orr, Roe, Reed, & Going, 2013).

During stance, forces are applied by the runner on the ground and vice-versa (ground reaction forces, GRFs). Mechanical work can be calculated as the dot product of GRFs and centre of mass (CoM) displacement (Cavagna & Kaneko, 1977). Similarly, mechanical power (the ratio of mechanical work and time) can be calculated as a dot product of GRFs and CoM velocity (Vigotsky, Zelik, Lake, & Hinrichs, 2019). The term "running power", however, may create confusion, since no net mechanical work, or power, is produced during the whole stance phase. Indeed, during the first half of the stance phase the CoM is being lowered in the vertical direction and decelerated in both vertical and horizontal (antero-posterior) directions, mechanical work and power are therefore negative. In the second half of the stance phase, the GRFs allow the runner to re-accelerate and lift the CoM. In this phase, mechanical work and power are positive. Thus, when running at constant velocity on flat ground, negative and positive mechanical work cancel each other out, leading to a power value of zero.

Various authors (Cavagna & Kaneko, 1977; Heglund, Fedak, Taylor, & Cavagna, 1982) have therefore proposed to consider only the positive increments of mechanical work and power for running. In addition to that, work and power due to the interaction of the CoM with the ground are typically referred to as "external" (Cavagna & Kaneko, 1977), as opposed to "internal", that takes in account the movement of the limbs in respect to the CoM. "Total work" is then defined as the sum of the "external" and "internal" work. However, a universally valid method to calculate mechanical work and power for running has not been established (Arampatzis, Knicker, Metzler, & Brüggemann, 2000), especially regarding "internal" work and power and the transfer, or the lack thereof, of kinetic energy among body segments (Cavagna & Kaneko, 1977; Kaneko, 1990; Williams & Cavanagh, 1983).

Mechanical work values can differ up to 100% based on different methods (Martin, Heise, & Morgan, 1993). In particular, Arampatzis et al. (2000) found that when mechanical power is calculated as the dot product of GRFs and velocity of CoM, that is: the "external" power according to Cavagna and Kaneko (1977), it revealed a significant linear relation with running velocity as opposed to kinematic methods that took in account also the "internal" contributions.

One commercially available device, Stryd (2017), uses an IMU and proprietary algorithms to estimate GRFs and CoM velocity changes in the vertical and horizontal directions, obtaining a running power that corresponds to the "external" power described by Arampatzis et al. (2000). Other devices, however, do not disclose how running power is calculated. In addition to running power, these devices provide other information to the runner, such as: contact time, aerial time, and step frequency (thanks to the use of IMUs), running speed (using GPS, global positioning system, data) and step length (combining step frequency and speed data). Other companies provide information about running power and biomechanics parameters by means of sensors embedded in the treadmill. All these parameters, in addition to running power, may be used by athletes and coaches to monitor training status and performance (Mohler, Fadillioglu, & Stein, 2021; Zrenner et al., 2020).

Aubry, Power, and Burr (2018) found only a weak relationship between running power calculated with the Stryd device and aerobic power. However, Snyder, Mohrman, Williamson, and Li (2018) evidenced methodological flaws in their analysis. A subsequent study by Cerezuela-Espejo et al. (2020) found a strong relationship between the running powers obtained by some of these wearable devices and aerobic power and proposed that in particular the Stryd device could be used as a surrogate of oxygen uptake ($\dot{V}O_2$) measurements. Imbach et al. (2020) found a significant effect of running running power, calculated using force-plates, and $\dot{V}O_2$. Similar to Cerezuela-Espejo et al. (2020), they also evidenced a linear relationship between running power measured by the Styd device and $\dot{V}O_2$ for each subject, however they report that the running power values obtained with the Stryd device were between 38% and 60% lower compared to those obtained using force-plates.

The objective of this paper is to evaluate the relationship between aerobic power and running powers measured using force-plates, and by different devices at different running speeds. Another objective of this paper is to compare the running powers and the various biomechanical parameters measured by different devices to force-plate measurements.

Methods

Experimental design

We enrolled a total of eleven subjects in the study (28.7 ± 6.6 years [range: 19–39 y]; 68.1 ± 12.0 kg [48–88 kg]; 1.75 ± 0.09 m [1.60–1.85 m]). They were active sport science students at the University of Udine, for four of them running was the main sport, while the remaining seven ran only occasionally. Subjects performed 5-min running trials at increasing speeds of 2.22, 2.78, 3.33, 3.89 and 4.44 m/s respectively (corresponding to 8, 10, 12, 14 and 16 km/h) with 5-min of recovery between trials. Each subject visited the lab once and performed all running trials while all metabolic and biomechanical measurements were collected simultaneously by different devices.

The experimental protocol was approved by the Ethics Committee of the University of Udine (Italy) and it was conducted according to the Declaration of Helsinki. Before the study, the purpose and objectives were carefully explained to each subject and written informed consent was obtained from all of them.

Aerobic power

We collected metabolic data by using a metabolic measurement system (CPET, Cosmed, Italy). Before each experimental session, we calibrated the volume and gas analysers using a 3-L calibration syringe and calibration gas (16.00% O_2; 4.00% CO_2), respectively. Then, we averaged the breath-by-breath $\dot{V}O_2$ and $\dot{V}CO_2$ values (in ml/s) during the last minute of each trial. We verified the reach of a steady state by observing an increase of no more than 1 ml/kg of O_2 throughout the last minute (Lazzer et al., 2015). From the $\dot{V}O_2$ and $\dot{V}CO_2$ measurements, we calculated aerobic power ($\dot{E}_{aero}$, in W) using the equation proposed by Brockway (1987):

$$\dot{E}_{aero} = 16.58 \times \dot{V}O_2 + 4.51 \times \dot{V}CO_2 \tag{1}$$

We then divided $\dot{E}_{aero}$ by each subject's body mass to obtain a normalised aerobic power ($\dot{E}_{aero,kg}$ in W/kg). As evidenced by Beck, Kipp, Byrnes, and Kram (2018), calculating $\dot{E}_{aero}$ using Equation (1) takes in account the ratio of oxidised carbohydrates to fats at different exercise intensities, as opposed to considering $\dot{V}O_2$ alone.

Among all subjects, the seven occasional runners started the protocol at the slowest speed (2.22 m/s),

Table 1. Biomechanical parameters obtained by force-based measurements and the other devices used in this study.

	Force-based			Technogym			Stryd			Garmin		
8 km/h ($n = 7$)												
Power (W/kg)	3.57	±	0.75	2.28	±	0.05	2.27	±	0.03	3.89	±	0.63
Contact time (s)	0.299	±	0.049	0.287	±	0.021	0.322	±	0.019	0.269	±	0.029
Aerial time (s)	0.074	±	0.036	0.089	±	0.014	0.056	±	0.008	0.105	±	0.020
Step frequency (Hz)	2.68	±	0.11	2.70	±	0.09	2.65	±	0.09	2.68	±	0.09
Step length (m)	0.83	±	0.03	0.84	±	0.03	0.76	±	0.02	0.90	±	0.15
Concentric phase (s)	0.221	±	0.085	0.200	±	0.019	–	–	–	–	–	–
Velocity (m/s)	2.22	±	0.0889	2.22	±	0.00	2.06	±	0.02	2.68	±	0.49
10 km/h ($n = 11$)												
Power (W/kg)	4.02	±	0.82	2.80	±	0.05	2.86	±	0.05	4.72	±	0.71
Contact time (s)	0.256	±	0.021	0.260	±	0.020	0.278	±	0.018	0.248	±	0.026
Aerial time (s)	0.105	±	0.017	0.103	±	0.013	0.088	±	0.009	0.114	±	0.023
Step frequency (Hz)	2.78	±	0.10	2.80	±	0.10	2.74	±	0.10	2.77	±	0.10
Step length (m)	1.00	±	0.04	1.01	±	0.03	0.96	±	0.04	1.14	±	0.17
Concentric phase (s)	0.172	±	0.05	0.179	±	0.01	–	–	–	–	–	–
Velocity (m/s)	2.78	±	0.00	2.78	±	0.00	2.60	±	0.05	3.36	±	0.52
12 km/h ($n = 11$)												
Power (W/kg)	4.64	±	0.74	3.39	±	0.22	3.42	±	0.05	5.10	±	0.84
Contact time (s)	0.236	±	0.022	0.243	±	0.018	0.248	±	0.014	0.233	±	0.028
Aerial time (s)	0.118	±	0.016	0.111	±	0.013	0.110	±	0.008	0.119	±	0.026
Step frequency (Hz)	2.84	±	0.10	2.87	±	0.10	2.80	±	0.09	2.84	±	0.10
Step length (m)	1.18	±	0.04	1.18	±	0.04	1.13	±	0.05	1.24	±	0.18
Concentric phase (s)	0.155	±	0.037	0.163	±	0.012	–	–	–	–	–	–
Velocity (m/s)	3.33	±	0.00	3.33	±	0.00	3.17	±	0.04	3.69	±	0.60
14 km/h ($n = 10$)												
Power (W/kg)	4.92	±	0.52	4.11	±	0.19	3.94	±	0.08	5.46	±	0.93
Contact time (s)	0.215	±	0.016	0.222	±	0.019	0.223	±	0.011	0.216	±	0.030
Aerial time (s)	0.127	±	0.014	0.122	±	0.013	0.126	±	0.010	0.126	±	0.029
Step frequency (Hz)	2.93	±	0.13	2.96	±	0.13	2.86	±	0.11	2.93	±	0.12
Step length (m)	1.33	±	0.06	1.34	±	0.06	1.30	±	0.07	1.34	±	0.19
Concentric phase (s)	0.127	±	0.023	0.147	±	0.011	–	–	–	–	–	–
Velocity (m/s)	3.89	±	0.00	3.89	±	0.00	3.71	±	0.07	4.12	±	0.66
16 km/h ($n = 3$)												
Power (W/kg)	5.53	±	0.24	4.86	±	0.21	4.53	±	0.03	5.47	±	0.85
Contact time (s)	0.208	±	0.007	0.202	±	0.006	0.206	±	0.006	0.201	±	0.007
Flight time (s)	0.124	±	0.006	0.138	±	0.006	0.139	±	0.006	0.137	±	0.019
Step frequency (Hz)	3.02	±	0.12	2.99	±	0.11	2.91	±	0.10	2.96	±	0.11
Step length (m)	1.48	±	0.06	1.52	±	0.05	1.51	±	0.03	1.43	±	0.08
Concentric phase (s)	0.123	±	0.002	0.13	±	0	–	–	–	–	–	–
Velocity (m/s)	4.44	±	0.00	4.44	±	0.00	4.30	±	0.04	4.23	±	0.34

Note: All values are mean ± standard deviation (SD). For each running speed, we report the number of subjects (n) included in our calculations, based on RER<1.0 (see methods for details).

while the remaining four started at 2.78 m/s. At the end of each 5-min trial, 5 min of recovery were enforced and the subsequent speed was selected based on each subject's feedback. In this analysis, we included only the completed trials in which the respiratory exchange ratio (RER) was below 1.0 for the whole duration of the last minute used to calculate $\dot{E}_{aero}$ (see Table 1).

Biomechanics

We collected GRFs of runners by mounting a treadmill (Skillrun, Technogym, Italy) on top of four force plates (BTS P-6000), in turn secured to the ground, in a configuration similar to what is described in Kram, Griffin, Donelan, and Chang (1998). GRFs were sampled at 1000 Hz, the vertical and horizontal (antero-posterior) components measured by each force plate were summed to obtain the total vertical and horizontal GRF. To measure the average speed of the treadmill during the running trials we used a tachometer (PLT200, Monarch Instrument, USA). We filtered the GRFs using a 4th order Butterworth filter with a 20 Hz cut-off (20). We calculated the instantaneous vertical velocity of the CoM by integrating the vertical acceleration of the CoM during stance (Cavagna, 1975).

Similarly, we calculated the instantaneous horizontal velocity of the CoM by integrating the horizontal acceleration of the CoM during stance and adding the average speed of the treadmill. We calculated the dot product vertical GRF (Figure 1(A)) and vertical velocity of the CoM to obtain instantaneous running power (in W) in the vertical direction and the dot product of horizontal GRF and horizontal velocity of the CoM to obtain instantaneous running power (in W) in the horizontal direction (Arampatzis et al., 2000). We then calculated the total instantaneous running power (P_{tot}, in W) as the sum of the vertical and horizontal powers at each instant (Figure 1(B)).

For each step, we identified the instant of touch down (t_{TD}) and toe off (t_{TO}) when the vertical component of the GRF exceeded a threshold of 10 N and decreased below the same threshold, respectively. In order to avoid "false" steps identifications (due to signal noise) for a step to be considered "valid" we verified that between t_{TD} and t_{TO} the vertical GRF exceeded 1 BW. Contact time (t_c) is defined as the duration between t_{TD} and t_{TO}, aerial time (t_a) as the duration between t_{TO} and t_{TD} of the following step (see Figure 1(A)). We then calculated step frequency (SF, In Hz) as:

$$SF = \frac{1}{t_c + t_a} \tag{2}$$

and step length (SL, in m) as:

$$SL = Speed \times (t_c + t_a) \tag{3}$$

where $Speed$ is the average speed of the treadmill in m/s measured with the digital encoder.

We identified at what instant t_0, during each stance, the total instantaneous running power switched from negative values to positive values. The duration of the eccentric phase (t_{ecc}, in s) is defined as:

$$t_{ecc} = (t_0 - t_{TD}) \tag{4}$$

and the duration of the concentric phase (t_{conc}, in s) is defined as:

$$t_{conc} = (t_{TO} - t_0) \tag{5}$$

See Figure 1(B). For each step, we calculated the average running power P_{avg} as:

$$P_{avg} = \frac{\int_{t_0}^{t_{TO}} P_{tot}\mathrm{d}t}{(t_c + t_a)} \tag{6}$$

where $\int_{t_0}^{t_{TO}} P_{tot}\mathrm{d}t$ is the time-integral of P_{tot} calculated throughout the concentric phase (that is: from t_0 to t_{TO}), and $(t_c + t_a)$ is the whole duration of the step.

For each trial (each subject/speed combination) we analysed the last 60 s, including only valid steps in our calculations. We then calculated the total average of each step P_{avg}. In order to compare different subjects, we divided said total average by each subject's body mass to obtain a normalised mechanical power ($P_{avg,kg}$, in W/kg).

While we collected GRF data from the forceplates, we also recorded the data provided by the following devices:

- Stryd (Stryd Summit Powermeter, firmware 1.2), a small (~ 0.01 kg) device attached to the shoe and paired to a sport watch (Garmin Fenix 3) worn on the right wrist.
- Garmin (Garmin HRM-Run, firmware 14.2), a ~ 0.06 kg device attached to a strap belt wore on the chest, similar to common heart-rate monitors, paired to a second sport watch (Garmin Fenix 5 plus) worn on the left wrist
- Technogym (Skillrun, Technogym, Italy), sensors measure the electric power absorbed and generated by the motor every time the foot makes contact with the treadmill belt

We calibrated each device before every test session as suggested by the corresponding manufacturer and we added personal data (age, weight, stature) to the devices before each test. We then imported the raw data (running power and respective biomechanical

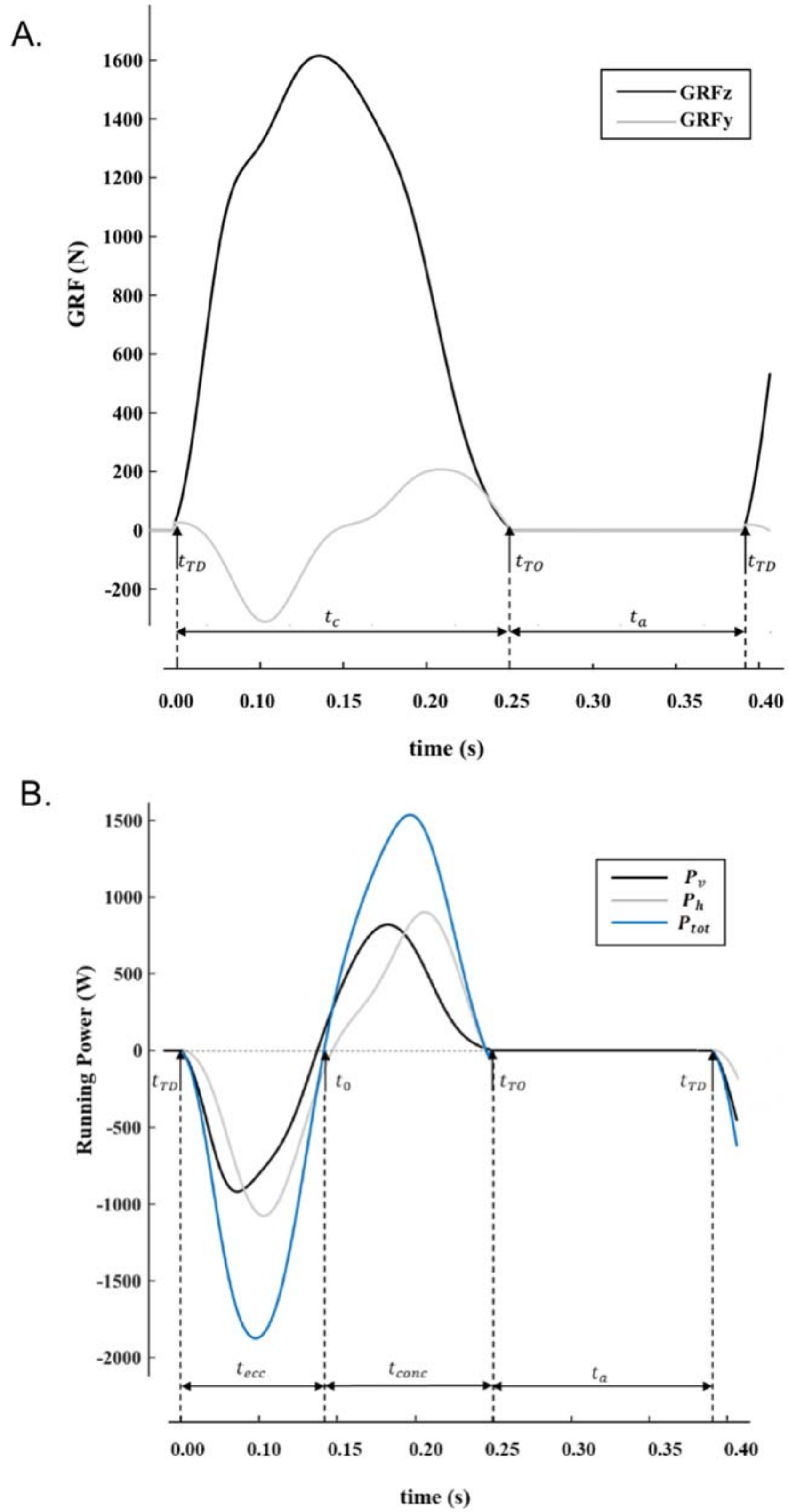

Figure 1. Raw data from a representative subject of vertical (GRFz, black line) and horizontal (GRFy, grey line) ground reaction forces (in N) as a function on time (panel A); and vertical (Pv, black line), horizontal (Ph, grey line) and total (Ptot, blue line) running power (in W) as a function on time (panel B). (t_{TD}: touch down; t_{TO}: toe off; t_c: contact time; t_a: aerial time; t_{ecc}: eccentric phase time; t_{conc}: concentric phase time).

parameters) from each portable device into MS Excel (Microsoft, USA) and for each subject, we calculated the average values of the last minute of each trial.

All tests were run on the same force-measuring treadmill setup described above and all metabolic and biomechanics data were collected simultaneously.

Statistical analyses

We used a linear mixed model to analyse the effects of running speed on aerobic power. We used a second linear mixed model to analyse the effects of running speed and the use of different devices (force-based, Technogym, Stryd, and Garmin) on running power and a third linear mixed model to analyse the relationship of aerobic power and the running power measured by different devices. Lastly, we use distinct linear mixed models to analyse the effects of speed and the use of different devices on the following biomechanical parameters: contact time, aerial time, concentric time, step length and step frequency. Each model was built using a step-down model-building approach (Kuznetsova, Brockhoff, & Christensen, 2017): first we included all the analysed factors and interactions, then the model was simplified until only statistically significant factors and interactions were present. For each model, we calculated the intraclass correlation coefficient (ICC): values of ICC>0.5 (Koo & Li, 2016), indicate that a linear mixed model was appropriate, as opposed to simple multiple linear regression analyses, to control for subject variability and to take in account the lack of independence between different measurements within the same subject. For models with values of ICC < 0.5, the subject random term was removed and, instead of a linear mixed model, a multiple linear regression model was calculated.

For multiple linear regression models we report multiple and adjusted R^2 values. For linear mixed models we calculated the conditional and marginal R^2 values according to Nakagawa, Johnson, and Schielzeth (2017).

We also used the Bland and Altman method (1986) to ascertain the bias between running power and other biomechanics measurements of different devices (Technogym, Stryd, and Garmin) compared to force-based measurements.

We used the R^2 values for the linear regressions calculated by Cerezuela-Espejo et al. (2020) between running power and $\dot{V}O_2$ for the Stryd and Garmin devices, to determine that for a multiple linear regression analysis with $\alpha = 0.05$, a statistical power of 0.80 and 4 different predictors (corresponding to the 4 devices compared in our paper), 11 subjects were enough.

We performed all statistical analyses using R-studio (RStudio Inc., Boston, MA) for each statistically significant relationship we report *p*-values (with significance set at $P < 0.05$), number of observations (N) and 95% confidence intervals (C.I.) for each statistically significant factor.

For each analysis, we implemented Bonferroni corrections to account for multiple comparisons.

Results

Effects of running speed on metabolic power and running power

We found a statistically significant linear correlation between metabolic power $\dot{E}_{aero}$ (in W/kg) and running speed (in m/s) (Figure 2(A)):

$$\dot{E}_{aero} = 1.505 + 4.056 \times Speed(\text{conditional } R^2 = 0.990, \text{ marginal } R^2 = 0.846, p < 0.001, N = 120) \tag{7}$$

95% C.I. were [0.621, 2.386] for the intercept and [3.960, 4.152] for the slope, respectively.

The ICC = 0.937 revealed a high correlation, indicating that $\dot{E}_{aero}$ measurements were clustered for each individual (that is: $\dot{E}_{aero}$ for each individual was consistently higher or lower compared to another individual across all speed range), justifying the use of a linear mixed model with subject random terms for Equation (7).

Running power (in W/kg) was affected by running speed (in m/s) and device (for the whole model: multiple $R^2 = 0.776$, adjusted $R^2 = 0.760$, $p < 0.001$ for both factors, $N = 160$), in particular:

$$Running\ Power_{Force} = 1.390 + 0.944 \times Speed \tag{8}$$

$$Running\ Power_{Garmin} = 1.877 + 0.944 \times Speed \tag{9}$$

$$Running\ Power_{Stryd} = 0.254 + 0.944 \times Speed \tag{10}$$

$$Running\ Power_{Technogym} = 0.296 + 0.944 \times Speed \tag{11}$$

95% C.I. were: [0.944, 1.837] for the intercept of Equation (8), [1.441, 2.313] for the intercept of Equation (9), [−0.188, 0.695] for the intercept of Equation (10), [−0.140, 0.732] for the intercept of Equation (11), respectively, and [0.818, 1.070] for the slope of all equations. Since ICC = 0.040 for the linear mixed model, Equations (8) to (11) are the result of the multiple linear regression model, without subject random terms.

Given that there was no statistically significant interaction between running speed and device ($p = 0.111$), following the step-down statistical approach described above, we fixed the value of slope ($\beta = 0.944$) associated with speed in all Equations (8) to (11). In other

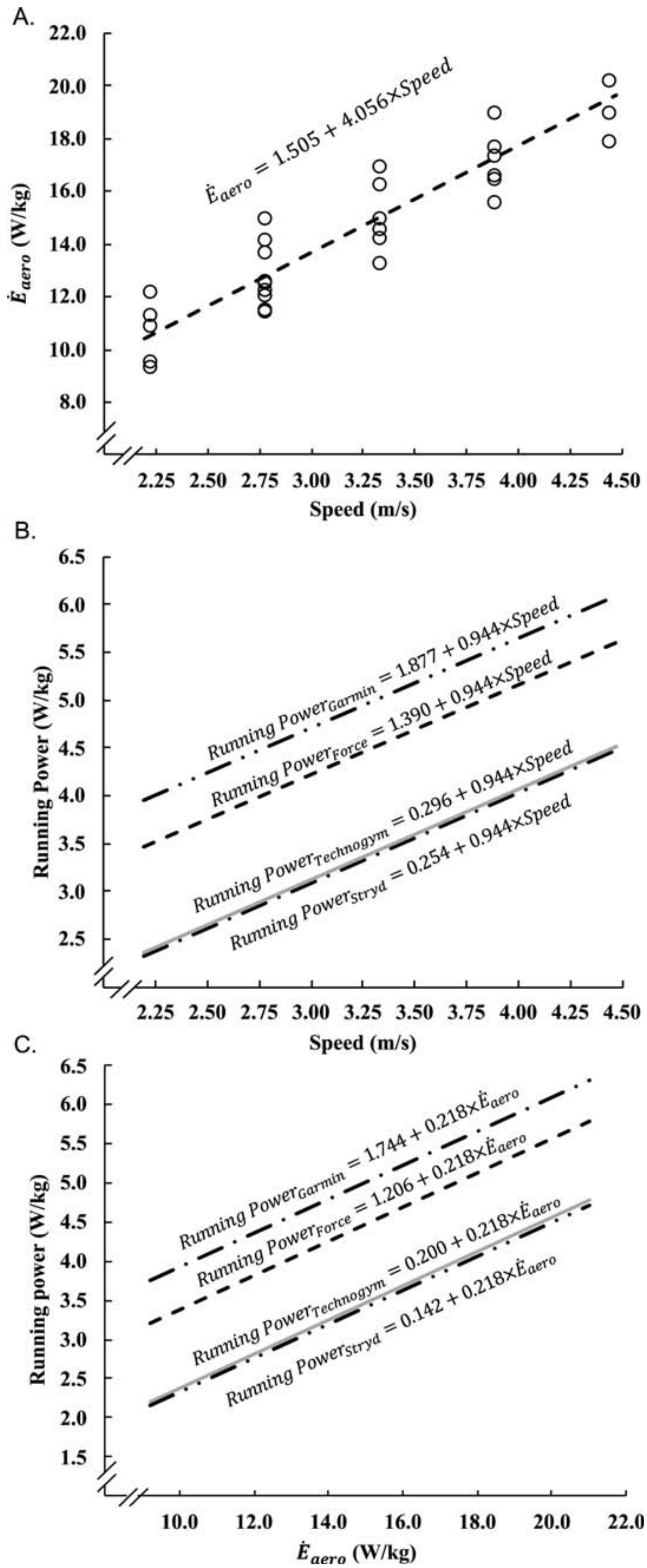

Figure 2. Metabolic power ($\dot{E}_{aero}$ in, W/kg, panel A) and running power (in W/kg, panel B) reported as a function of treadmill speed (in m/s). Running power as a function of $\dot{E}_{aero}$ for different devices (panel C).

Table 2. Bland-Altman comparisons between Force-based measurements and the other devices used in this study.

	Technogym					Stryd					Garmin				
				95% Limit of agreement					95% Limit of agreement					95% Limit of agreement	
	BIAS		SD	From	to	BIAS		SD	From	to	BIAS		SD	From	to
Power (W/kg)	−1.066	±	0.705	−2.449	0.316	−1.044	±	0.627	−2.273	0.185	0.513	±	1.176	−1.793	2.818
Contact time (s)	0.002	±	0.014	−0.026	0.029	0.011	±	0.016	−0.021	0.043	−0.010	±	0.023	−0.055	0.036
Aerial time (s)	0.001	±	0.013	−0.026	0.027	−0.006	±	0.017	−0.039	0.028	0.010	±	0.023	−0.036	0.055
Step frequency (Hz)	0.027	±	0.032	−0.035	0.089	−0.044	±	0.040	−0.122	0.033	−0.002	±	0.028	−0.058	0.054
Step length (m)	0.007	±	0.015	−0.021	0.036	−0.040	±	0.028	−0.095	0.016	0.061	±	0.181	−0.294	0.415
Concentric phase (s)	0.007	±	0.039	−0.069	0.082	--		--	--	--	--		--	--	--
Velocity (m/s)	0.000	±	0.000	0.000	0.000	−0.168	±	0.052	−0.271	−0.066	0.363	±	0.577	−0.768	1.494

words: an increase in running speed leads to an identical increase in running power measured by each device, and only the difference in intercept was statistically significant (see Appendix). In particular, across running speeds, running power was 35% higher for Garmin, 81% lower for Stryd and 78% lower for Technogym compared to force-based ($p < 0.001$ for all comparisons, Figure 2(B)).

The Bland & Altman (5) plots reflect similar differences across the range of measured running powers (Table 2). The mean differences for Technogym and Stryd are −1.07 W/kg and −1.04 W/kg, respectively, evidencing a negative bias in running power compared to force-based measurements. Garmin, on the other hand, has a positive bias of +0.51 W/kg compared to force-based measurements. Similarly to the linear mixed model analysis, where we found no interaction between running speed and device, we did not find a trend in the differences, indicating a lack of proportionality between the calculated differences and the average running power (see also Table 1).

Relationships between aerobic power and running power

We found a statistically significant linear correlation between aerobic power ($\dot{E}_{aero}$, in W/kg) and running power (in W/kg) measured by the different devices (for the whole model: multiple $R^2 = 0.787$, adjusted $R^2 = 0.748$, $p < 0.001$ for both factors, $N = 118$). In particular:

$$Running\ Power_{Lab} = 1.206 + 0.218 \times \dot{E}_{aero} \quad (12)$$

$$Running\ Power_{Garmin} = 1.744 + 0.218 \times \dot{E}_{aero} \quad (13)$$

$$Running\ Power_{Stryd} = 0.142 + 0.218 \times \dot{E}_{aero} \quad (14)$$

$$Running\ Power_{Technogym} = 0.200 + 0.218 \times \dot{E}_{aero} \quad (15)$$

95% C.I. were: [0.665, 1.748] for the intercept of Equation (12), [1.211, 2.277] for the intercept of Equation (13), [−0.391, 0.675] for the intercept of Equation (14), [−0.333, 0.733] for the intercept of Equation (15), respectively, and [0.184, 0.253] for the slope of all equations. Since ICC = 0.119 for the linear mixed model, Equations (12) to (15) are the result of the multiple linear regression model, without subject random terms.

Given that there was no statistically significant interaction between aerobic power and devices ($p = 0.350$), we fixed the value of slope ($\beta = 0.218$) associated with $\dot{E}_{aero}$ in all Equations (12) to (15). In other words: an increase in aerobic power leads to an identical increase in running power measured by each device, and only the difference in intercept was statistically significant. In particular, across aerobic powers, running power was 46% higher for Garmin, 88% lower for Stryd and 83% lower for Technogym compared to force-based ($p < 0.001$ for all comparisons, Figure 2(C)).

Biomechanical parameters

We found a statistically significant linear correlation between running speed, in m/s, and contact times, in s, measured by different devices, with a statistically significant interaction between running speed and device factors (for the whole model: conditional $R^2 = 0.884$, marginal $R^2 = 0.636$, $p < 0.001$, $N = 160$). In particular:

$$t_{cForce} = 0.385 - 0.043 \times Speed \quad (16)$$

$$t_{cGarmin} = 0.333 - 0.030 \times Speed \quad (17)$$

$$t_{cStryd} = 0.425 - 0.052 \times Speed \quad (18)$$

$$t_{cTechnogym} = 0.367 - 0.037 \times Speed \quad (19)$$

95% C.I. were [0.362, 0.408] for the intercept and [−0.049, −0.037] for the slope of Equation (16), [0.312, 0.354] for the intercept and [−0.038, −0.022] for the slope of Equation (17), [0.403, 0.447] for the intercept and [−0.061, −0.044] for the slope of Equation (18), [0.345, 0.388] for the intercept and [−0.045, −0.029] for the slope of Equation (19), respectively. The ICC = 0.682 revealed a high correlation, justifying the use of a linear mixed model with subject random terms for Equation (16) to (19).

Compared to force-based, the intercept values were 14% lower for Garmin ($p < 0.001$), 10% higher for Stryd ($p = 0.004$), there was no statistically significant difference for Technogym ($p = 0.183$). Compared to force-based the slopes were 30% higher for Garmin ($p = 0.001$), 21% lower for Stryd ($p = 0.033$), there was no statistically significant difference for Technogym ($p = 0.146$). The Bland & Altman analysis evidenced how across the range of measured contact times, mean differences were −0.010 s for Garmin, +0.011 s for Stryd and +0.002 s for Technogym, respectively (Table 2).

We found a statistically significant correlation between running speed, in m/s, and concentric times, in s, (multiple $R^2 = 0.379$, adjusted $R^2 = 0.371$, $p < 0.001$, $N = 80$), but no statistically significant difference ($p = 0.333$) between force-based and Technogym (note that neither Stryd or Garmin devices provide this measurement):

$$t_{conc} = 0.285 - 0.038 \times Speed \quad (20)$$

95% C.I. were: [0.249, 0.321] for the intercept and [−0.048, −0.027] for the slope, respectively. Since ICC = 0.434 for the linear mixed model, Equation (20) is the result of the multiple linear regression model, without subject random terms. The Bland & Altman analysis evidenced how across the range of measured concentric times, the mean difference was +0.007 s for Technogym compared to force-based measurements (Table 2).

We found no statistically significant correlation between running speed and aerial times measured by different devices ($p = 0.059$). The Bland & Altman analysis evidenced how across the range of measured aerial times, mean differences were +0.010 s for Garmin, −0.006 s for Stryd and +0.001 s for Technogym, respectively (Table 2).

We found a statistically significant correlation between running speed, in m/s, and step length, in m, measured by different devices, with a statistically significant interaction between running speed and device factors (for the whole model: multiple $R^2 = 0.785$, adjusted $R^2 = 0.779$, $p < 0.001$, $N = 146$), in particular:

$$SL_{Force} = 0.222 + 0.284 \times Speed \quad (21)$$

$$SL_{Garmin} = 0.281 + 0.284 \times Speed \quad (22)$$

$$SL_{Stryd} = 0.181 + 0.284 \times Speed \quad (23)$$

$$SL_{Technogym} = 0.229 + 0.284 \times Speed \quad (24)$$

95% C.I. were: [0.134, 0.310] for the intercept of Equation (21), [0.195, 0.367] for the intercept of Equation (22), [0.090, 0.272] for the intercept of Equation (23), [0.143, 0.315] for the intercept of Equation (24), and [0.259, 0.309] for the slopes of all equations. Since ICC = 0.250 for the linear mixed model, Equations (21) to (24) are the result of the multiple linear regression model, without subject random terms.

Given that there was no statistically significant interaction between speed and devices ($p = 0.059$), we fixed the value of slope ($\beta = 0.284$) associated with *Speed*. Compared to force-based the intercept was 26% higher for Garmin ($p = 0.010$), there were no statistically significant differences for Stryd ($p = 0.113$) and Technogym ($p = 0.752$). The Bland & Altman analysis evidenced how across the range of measured step lengths, mean differences were +0.061 m for Garmin, −0.040 m for Stryd and +0.007 m for Technogym, respectively (Table 2).

We found a statistically significant correlation between running speed, in m/s, and step frequency, in Hz, measured by different devices with a statistically significant interaction between running speed and device factors (for the whole model: conditional $R^2 = 0.955$, marginal $R^2 = 0.515$, $p < 0.001$, $N = 160$), in particular:

$$SF_{Force} = 2.327 + 0.156 \times Speed \quad (25)$$

$$SF_{Garmin} = 2.358 + 0.145 \times Speed \quad (26)$$

$$SF_{Stryd} = 2.380 + 0.126 \times Speed \quad (27)$$

$$SF_{Technogym} = 2.372 + 0.150 \times Speed \quad (28)$$

95% C.I. were: [2.252, 2.402] and [0.141, 0.171] for the intercept and slope of Equation (25), [2.285, 2.430] and [0.126, 0.165] for the intercept and slope of Equation (26), [2.306, 2.454] and [0.106, 0.146] for the intercept and slope of Equation (27), [2.299, 2.445] and [0.130, 0.169] for the intercept and slope of Equation (28). The ICC = 0.907 revealed a high correlation, justifying the use of a linear mixed model with subject random terms for Equation (25) to (28).

Compared to force-based the intercepts were not statistically significant different for Garmin ($p = 0.356$), Stryd ($p = 0.121$) and Technogym ($p = 0.176$). Compared to lab based, the slope was 19% lower for Stryd ($p = 0.004$), there were no statistically significant differences for Garmin ($p = 0.290$) and Technogym ($p = 0.535$). The Bland & Altman analysis evidenced how across the range of measured step frequencies, mean differences were −0.002 Hz for Garmin, −0.044 Hz for Stryd and +0.027 Hz for Technogym, respectively (Table 2).

Discussion

Not surprisingly, aerobic power measurements are in accordance with previous literature (4, 22), evidencing how metabolic demand increases linearly with running speed at sub-maximal intensities. Batliner et al. (2018) found that, when using a linear relationship between running velocity and aerobic power, the slope was 4.046 and 4.256 for recreational and sub-elite runners respectively, matching the value of 4.056 reported in the present study. Conversely, the intercepts (0.588 and −0.767, for recreational and sub elite runners respectively) seem to be 61% and 151% lower compared to the value of 1.505 reported in this study. According to Equation (7), the intercept represents the metabolic power needed when running speed equals 0 m/s, i.e. the resting metabolic rate: 1.505 W/kg in fact corresponds to ~4 mlO2/min/kg (Hoogkamer, Taboga, & Kram, 2014). While a negative value has no physiological meaning, different (positive) values of the intercept could be explained by different resting metabolic rates of the subjects. However, caution should be used when trying to extrapolate metabolic values outside the range of running velocities used to calculate these models. For example, while Batliner et al. (2018) evidenced how a curvilinear model is better suited to determine aerobic powers at faster running velocities compared to a linear model, at 0 m/s their curvilinear models would predict extremely high resting metabolic rates: 3.349 and 5.721 W/kg for recreational and sub elite runners respectively.

Force-based running power measurements reported in this study support those described by Cavagna, Thys, and Zamboni (1976) in the same range of velocities: with values ranging from 3.2 W/kg to 6.0 W/kg when speed increases from 2.22 m/s to 4.44 m/s. The values reported in Arampatzis et al. (2000) are much higher compared to the present study and to previous literature, ranging from 9.4 W/kg at approx. 2.5 m/s to 19.2 W/kg at approx. 4.5 m/s. However, Arampatzis et al. (2000) calculated running power as the ratio between external work and contact time, differing from our study where we calculate the average running power throughout the whole duration of a step (the sum of contact and aerial time, see Equation (6)). Dividing the same amount of external work by t_c alone, instead of $(t_c + t_a)$, leads to higher values of power. This discrepancy in the gait times used when calculating average values could therefore explain the apparently higher running powers reported by Arampatzis et al. (2000).

We found a statistically significant difference between force-based running power and the running power measured by other devices (see Equations (8) to (11)). All intercept values are positive, evidencing the fact that according to the linear model even when running speed is zero, the running power is positive for force-based measurements and all other devices. This could be explained at least partially by the fact that if a subject is running "in place", that is: bouncing up and down with no forward displacement, they would still accelerate in the vertical direction during the concentric phase, therefore producing a positive running power. At running speeds faster than zero, this vertical motion is accompanied by horizontal deceleration and acceleration. As speed increases, the vertical displacement of the centre of mass during contact decreases (Farley & Ferris, 1998), this is accompanied by higher ground reaction forces, meanwhile, the contribution of horizontal deceleration and acceleration increases. As evidenced by Cavagna, Heglund, and Willems (2005) and confirmed in our study, this leads to an overall linear increase in running power as a function of running speed. As evidenced by Equations (8) to (11), the linear mixed models evidenced no difference in the slope for force-based and the other devices, but the difference in intercept values was statistically significant. For the calculation of force-based instantaneous total power, we calculated the powers in the vertical and horizontal directions, total instantaneous power was then calculated by summing those powers at every instant (Figure 1(B)) and then averaged throughout the whole step. This same method is being used by the Stryd device (2017). Given the similarities between the power measurements with the Stryd device (see Figure 2) we can reasonably assume that a similar approach was used by Technogym device, however, this can't be verified since the company did not disclose their algorithms. For example, both Technogym and Garmin devices may use slightly different approaches (for example calculating "vertical" and "horizontal" powers throughout the step and then summing their average values, or assign different weights to the power produced in the vertical and horizontal directions), explaining the observed differences in Equations (8) to (11). However, as evidenced by Vigotsky et al. (2019) these "directional" powers do not necessarily represent the rate of work performed in a given direction and caution should be used when assigning vector quantities to powers. In particular, if power is averaged only in the concentric phase of the step, instead of the whole duration, this could lead to values that are basically twice as big as the powers calculated in Equation (6) of this manuscript, possibly explaining the consistently elevated measurements obtained with the Garmin device compared to the other devices. However, we would like to point out that speculating on the exact algorithms used by these devices when calculating running power is beyond the scope of our paper. In addition to this positive bias, the limits of agreement in the Bland–Altman plot comparing the Garmin device to force-plate measurements ([−1.793,2.818]) are larger compared to those of the Technogym and Stryd devices ([−2.449,0.316] and

[−2.273,0.185], respectively), indicating a higher variability (a lower precision) of this device when measuring running power compared to force-based measurements.

We found that running power increases linearly with aerobic power. Given that the only difference between force-based running power measurements and the other devices was the intercept, when expressed as a function of running speed, it is not surprising that similar differences are evidenced when expressing running power as a function of aerobic power. Our results support the conclusions of Cerezuela-Espejo et al. (2020) and Imbach et al. (2020), who found that running power increases linearly with aerobic power. Unfortunately, both Cerezuela-Espejo et al. (2020) and Imbach et al. (2020) do not report aerobic power in W/kg but only oxygen consumption, therefore an accurate conversion into W/kg is not possible (Beck et al., 2018), and running power is reported in W (not in W/kg) so we can't make direct comparison with the values presented in our study. Notwithstanding the differences in intercept, Equations (12) to (15) evidence how an increase of aerobic power of 1 W/kg translates directly to an increase of running power of 0.218 W/kg for all devices and force-based measurements. While we did not test this directly, runners could use running power measurements to monitor their aerobic fitness during different training phases: for a given running speed, a decrease in running power translates directly to a decrease in aerobic power, indicating an improvement in running economy. Running power is independent from internal (physiological and/or psychological) factors such as fatigue, stress, etc. or external factors such as temperature and humidity, providing an additional useful metric to monitor training load compared to common methods such heart rate, running pace/duration etc. (Paquette, Napier, Willy, & Stellingwerff, 2020). In addition, as suggested by some manufacturers, athletes can perform critical power tests and set specific training zones based on running power: as opposed to heart rate, with an intrinsic latency time, running power is calculated basically instantaneously and can provide quasi real-time information during training sessions.

As evidenced in Tables 1 and 2, the reported running speed is identical between force-based and Technogym: as explained in the methods, the Technogym treadmill was mounted on top of the force plates and its speed was selected as the running speed of each subject. On the other hand, given that GPS data can't be used on a treadmill, Stryd and Garmin devices calculate speed likely based on step length and step frequency measurements. Stryd underestimates both step length (bias = −0.040 m) and step frequency (bias = −0.044 Hz), it is therefore not surprising that the reported speed is also slower (bias = −0.168 m/s). The Garmin device, while reporting a substantially identical step frequency as the force-based measurement (bias = −0.002 Hz), overestimates step length (bias = +0.060 m), this in turn translates in a faster velocity (bias = +0.363 m/s).

Limitations

In this paper, as suggested by various authors (Cavagna & Kaneko, 1977; Heglund et al., 1982), we report running power as the positive portion of mechanical power (in other words, we focus only on the concentric phase). The Stryd device uses the same methodology (Snyder, Kipp, & Hoogkamer, 2021) and we can reasonably assume that the other devices use a similar approach. We must however keep in mind that the true **net** mechanical power produced when running on flat terrain at constant speed, notwithstanding the power needed to overcome air resistance, is zero: at the end of each step the centre of mass of a runner has the same potential and kinetic energy it had at the end of the previous step. This would lead to efficiency values of zero (Alexander, 1991). Running power measurements are therefore highly dependent on how eccentric and concentric phases are identified, in addition to how accurately vertical and horizontal oscillation of the centre of mass is calculated, and also on *if* and *how* internal power is included in the computation.

We need to emphasise that all of our measurements were performed indoor on a level treadmill. Future studies should investigate the effects of different uphill/downhill gradients, the effect of head/tail wind and different terrains (concrete, grass, etc.), to generalise the validity of running power metrics in different outdoor settings.

In conclusion: we report that running power increases linearly with running speed. While there are differences between devices and force-based running power measurements, increasing running speed translates to identical increases in external running power, irrespective of device or force-based measurements. Similarly, increases in aerobic power translates to identical increases in running power, irrespective of device or force-based measurements. Therefore, running power measured by commercially available devices reflects force-based measurements and it can be a valuable metric to evaluate training sessions and competitions.

Acknowledgements

The authors thank the subjects who agreed to participate to the study. We thank Kristine L. Snyder Ph.D. for her meticulous help with verifying our calculations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Technogym S.p.A (Cesena, FC, Italy); Fondazione P. Pittini (Italy).

ORCID

Paolo Taboga http://orcid.org/0000-0001-6529-8299
Nicola Giovanelli http://orcid.org/0000-0002-6914-2117
Stefano Lazzer http://orcid.org/0000-0003-0067-2221

References

Stryd. (2017). Retrieved December 6, 2020, from https://storage.googleapis.com/stryd_static_assets/stryd-metric-validation.pdf.

Alexander, R. M. (1991). Energy-saving mechanisms in walking and running. *Journal of Experimental Biology*, *160*, 55–69.

Arampatzis, A., Knicker, A., Metzler, V., & Brüggemann, G.-P. (2000). Mechanical power in running: A comparison of different approaches. *Journal of Biomechanics*, *33*(4), 457–463.

Aubry, R. L., Power, G. A., & Burr, J. F. (2018). An assessment of running power as a training metric for elite and recreational runners. *Journal of Strength and Conditioning Research*, *32*(8), 2258–2264.

Batliner, M. E., Fujii, A., Fujita, M., Fujiwara, R., Emoto, K., Sano, K., … Yagi, H. (2018). Does metabolic rate increase linearly with running speed in all distance runners? *Sports Medicine International Open*, *2*(1), E1–E8.

Beck, O. N., Kipp, S., Byrnes, W. C., & Kram, R. (2018). Use aerobic energy expenditure instead of oxygen uptake to quantify exercise intensity and predict endurance performance. *Journal of Applied Physiology*, *125*(2), 672–674.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310.

Brockway, J. M. (1987). Derivation of formulae used to calculate energy expenditure in man. *Human Nutrition. Clinical Nutrition*, *41*(6), 463–471.

Capostagno, B., Lambert, M. I., & Lamberts, R. P. (2016). A systematic review of submaximal cycle tests to predict, monitor, and optimize cycling performance. *International Journal of Sports Physiology and Performance*, *11*(6), 707–714.

Cavagna, G. A. (1975). Force platforms as ergometers. *Journal of Applied Physiology*, *39*(1), 174–179.

Cavagna, G. A., Heglund, N. C., & Willems, P. A. (2005). Effect of an increase in gravity on the power output and the rebound of the body in human running. *Journal of Experimental Biology*, *208*(Pt 12), 2333–2346.

Cavagna, G. A., & Kaneko, M. (1977). Mechanical work and efficiency in level walking and running. *The Journal of Physiology*, *268*, 467–481.

Cavagna, G. A., Thys, H., & Zamboni, A. (1976). The sources of external work in level walking and running. *The Journal of Physiology*, *262*(3), 639–657.

Cerezuela-Espejo, V., et al. (2020). Are we ready to measure running power? Repeatability and concurrent validity of five commercial technologies. *European Journal of Sport Science*, *21*(3), 341–350.

Farley, C. T., & Ferris, D. P. (1998). Biomechanics of walking and running: Center of mass movements to muscle action. *Exercise and Sport Sciences Reviews*, *26*, 253–285.

Heglund, N. C., Fedak, M. A., Taylor, C. R., & Cavagna, G. A. (1982). Energetics and mechanics of terrestrial locomotion. IV. Total mechanical energy changes as a function of speed and body size in birds and mammals. *Journal of Experimental Biology*, *97*, 57–66.

Hongu, N., Orr, B. J., Roe, D. J., Reed, R. G., & Going, S. B. (2013). Global positioning system watches for estimating energy expenditure. *Journal of Strength and Conditioning Research*, *27*(11), 3216–3220.

Hoogkamer, W., Taboga, P., & Kram, R. (2014). Applying the cost of generating force hypothesis to uphill running. *PeerJ*, *2*, e482.

Imbach, F., et al. (2020). Validity of the Stryd power meter in measuring running parameters at submaximal speeds. *Sports (Basel)*, *8*(7), 103–121.

Jaen-Carrillo, D., Roche-Seruendo, L. E., Cartón-Llorente, A., Ramírez-Campillo, R., & García-Pinillos, F. (2020). Mechanical power in endurance running: A scoping review on sensors for power output estimation during running. *Sensors (Basel)*, *20*(22), 6482.

Kaneko, M. (1990). Mechanics and energetics in running with special reference to efficiency. *Journal of Biomechanics*, *23* (Suppl 1), 57–63.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Kram, R., Griffin, T. M., Donelan, J. M., & Chang, Y. H. (1998). Force treadmill for measuring vertical and horizontal ground reaction forces. *Journal of Applied Physiology*, *85*(2), 764–769.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lazzer, S., Salvadego, D., Taboga, P., Rejc, E., Giovanelli, N., & di Prampero, P. E. (2015). Effects of the Etna uphill ultramarathon on energy cost and mechanics of running. *International Journal of Sports Physiology and Performance*, *10*(2), 238–247.

Martin, P. E., Heise, G. D., & Morgan, D. W. (1993). Interrelationships between mechanical power, energy transfers, and walking and running economy. *Medicine & Science in Sports & Exercise*, *25*(4), 508–515.

Mohler, F., Fadillioglu, C., & Stein, T. (2021). Fatigue-related changes in spatiotemporal parameters, joint kinematics and leg stiffness in expert runners during a middle-distance run. *Frontiers in Sports and Active Living*, *3*, 634258.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R(2) and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, *14*(134), 20170213.

Paquette, M. R., Napier, C., Willy, R. W., & Stellingwerff, T. (2020). Moving beyond weekly 'distance': Optimizing quantification of training load in runners. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(10), 564–569.

Paton, C. D., & Hopkins, W. G. (2001). Tests of cycling performance. *Sports Medicine*, *31*(7), 489–496.

Rodriguez, D. A., Brown, A. L., & Troped, P. J. (2005). Portable global positioning units to complement accelerometry-based physical activity monitors. *Medicine & Science in Sports & Exercise*, *37*(11 Suppl), S572–S581.

Snyder, K. L., Kipp, S., & Hoogkamer, W. (2021). *Running power definition and utility*. https://www.stryd.com/running-power-definition.

Snyder, K. L., Mohrman, W. P., Williamson, J. A., & Li, K. (2018). Methodological flaws in Aubry, RL, Power, GA, and Burr, JF. An assessment of running power as a training metric for elite and recreational runners. J Strength Cond Res 32: 2258–2264, 2018. *Journal of Strength and Conditioning Research*, *32*(12), e61.

Townshend, A. D., Worringham, C. J., & Stewart, I. B. (2008). Assessment of speed and position during human locomotion using nondifferential GPS. *Medicine & Science in Sports & Exercise*, *40*(1), 124–132.

Vigotsky, A. D., Zelik, K. E., Lake, J., & Hinrichs, R. N. (2019). Mechanical misconceptions: Have we lost the “mechanics” in “sports biomechanics”? *Journal of Biomechanics*, *93*, 1–5.

Williams, K. R., & Cavanagh, P. R. (1983). A model for the calculation of mechanical power during distance running. *Journal of Biomechanics*, *16*(2), 115–128.

Zrenner, M., Küderle, A., Roth, N., Jensen, U., Dümler, B., & Eskofier, B. M. (2020). Does the position of foot-mounted IMU sensors influence the accuracy of spatio-temporal parameters in endurance running? *Sensors (Basel)*, *20*(19), 5705.

Appendix. Step by step description of our statistical approach

Equations (8)–(10)

Following (Kuznetsova et al., 2017), in the first iteration of our analysis, we estimated the effects of *Speed* (continuous variable) and *Device* (a categorical variable, indicated by *factor* in R-Studio) on *Power*, including the interaction of the two variables (*Speed* Device)*

*RunPowervsSpeed1 <-lm(PowerKg~Speed*factor(Device), data=Total)*

We obtained the following result:

*lm(formula = PowerKg ~ Speed * factor(Device), data = Total)*

Residuals:

Min	*1Q*	*Median*	*3Q*	*Max*
−1.3595	−0.1243	−0.0014	0.1607	1.4026

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.6920	0.4403	3.842	0.000179 ***
Speed	0.8515	0.1319	6.455	1.37e-09 ***
factor(Device)GARMIN	0.7370	0.5979	1.233	0.219561
factor(Device)STRYD	−1.6333	0.6129	−2.665	0.008531 **
factor(Device)TECHNOGYM	−2.0303	0.5979	−3.396	0.000873 ***
Speed:factor(Device)GARMIN	−0.0796	0.1805	−0.441	0.659884
Speed:factor(Device)STRYD	0.1525	0.1847	0.826	0.410270
Speed:factor(Device) TECHNOGYM	0.2895	0.1805	1.603	0.110914

*Signif. codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “ ” 1Residual standard error: 0.5263 on 152 degrees of freedom(8 observations deleted due to missingness)Multiple R-squared: 0.7735, Adjusted R-squared: 0.7631F-statistic: 74.15 on 7 and 152 DF, p-value: < 2.2e-16*

As evidenced by the last 3 lines: all interactions between *Speed* and *Device* were above the threshold of 0.05 (the smallest value being 0.11), meaning that *Speed* and *Device* variables affect *Power*, but independently. The final model is then re-calculated removing this non-statistically significant interaction:

RunPowervsSpeed2 <-lm(Power~Speed + factor(Device) data=Total)

Obtaining the following results:

Call:

lm(formula = PowerKg ~ Speed + factor(Device), data = Total)

Residuals:

Min	*1Q*	*Median*	*3Q*	*Max*
−1.57074	−0.15813	−0.02234	0.16626	1.40724

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.39021	0.22602	6.151	6.28e-09 ***
Speed	0.94361	0.06383	14.78	<2e-16 ***
factor(Device)GARMIN	0.48676	0.11873	4.100	6.66e-05 ***
factor(Device)STRYD	−1.13655	0.12163	−9.344	<2e-16 ***
factor(Device) TECHNOGYM	−1.09425	0.11873	−9.216	<2e-16 ***

*Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1Residual standard error: 0.53 on 155 degrees of freedom (8 observations deleted due to missingness)Multiple R-squared: 0.7657, Adjusted R-squared: 0.7597F-statistic: 126.7 on 4 and 155 DF, p-value: < 2.2e-16*The Estimate values were then included in Equations (8)–(11). (Intercept) corresponds to the Lab measurement, our reference; to obtain the equations for the other devices, the value of the corresponding Estimate was added to the reference value.