Journal of Medical Engineering & Technology

ISSN: 0309-1902 (Print) 1464-522X (Online) Journal homepage: http://www.tandfonline.com/loi/ijmt20

Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities

Jomme Claes , Roselien Buys, Andrea Avila, Dewar Finlay, Alan Kennedy, Daniel Guldenring, Werner Budts & Véronique Cornelissen

To cite this article: Jomme Claes , Roselien Buys, Andrea Avila, Dewar Finlay, Alan Kennedy, Daniel Guldenring, Werner Budts & Véronique Cornelissen (2017): Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities, Journal of Medical Engineering & Technology, DOI: 10.1080/03091902.2017.1333166

To link to this article: http://dx.doi.org/10.1080/03091902.2017.1333166

Published online: 04 Jul 2017.

JOURNAL OF MEDICAL ENGINEERING & TECHNOLOGY, 2017
https://doi.org/10.1080/03091902.2017.1333166

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities

Jomme Claes[a], Roselien Buys[b], Andrea Avila[b], Dewar Finlay[c], Alan Kennedy[c], Daniel Guldenring[c], Werner Budts[a] and Véronique Cornelissen[b]

[a]Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium; [b]Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium; [c]University of Ulster, Belfast, UK

ABSTRACT

The accuracy of wrist worn heart rate monitors based on photoplethysmography (PPG) is not fully clinically accepted. Therefore, we aimed to validate heart rate measurements of a commercially available PPG heart rate monitor, i.e. the Garmin Forerunner® 225. Twelve healthy volunteers (six women; mean age: 28 years) performed a treadmill protocol consisting of: five minutes sitting, five minutes standing, 10 minutes walking at 4 km/h, 10 minutes walking at a gradient of 5% and intensity of 4–6 metabolic equivalents (METs), 10 minutes walking at a gradient of 8% and intensity of seven METs or more. Walking speeds were individually determined. Walking bouts were separated by a standardised five minute rest period. Heart rate was measured as the average of the last three minutes standing and of each walking bout. A three lead patch-based electrocardiogram (ECG; Zensor®) was used as criterion method. Statistical analyses included Pearson's correlation (r), paired t-tests, root mean squared error (RMSE) and Bland–Altman plots. The mean values per three minutes of every condition did not differ significantly between the Garmin Forerunner® 225 and the Zensor®. RMSE was 3.01 beats per minute (bpm) or 2.89%. The Bland–Altman bias was 1.57 bpm. Limits of agreement (LoA) were wide, ranging from 32.53 to 29.40 bpm. However, Pearson's r ranged from 0.650 to 0.868 suggesting moderate to strong validity. Generally, mean heart rates, r values, RMSE and the Bland–Altman bias indicated good overall agreement in this sample of healthy adults, but wide LoA are making it difficult to trust individual measurements.

ARTICLE HISTORY
Received 24 February 2017
Revised 12 May 2017
Accepted 14 May 2017
Published online 3 July 2017

KEYWORDS
Validity; Garmin; heart rate; photoplethysmography; sports watch

1. Introduction

Wearable technologies to measure health and performance are trending. A recent survey stated that wearable technologies will top the list of fitness trends in 2016 for the first time [1]. This is in line with the observation that wearable heart rate monitors are increasingly used to guide not only the athletes but also healthy individuals and patients with chronic diseases while exercising. Exercising at the optimal intensity is crucial to improve performance and cardiorespiratory fitness [2]. The validity of the earliest versions of heart rate monitors, involving a wireless chest strap that interfaces with a digital watch, has been abundantly demonstrated [3,4]. The newer generation of heart rate monitors use photoplethysmography (PPG) with optical sensors at the level of the wrist, to continuously measure heart rate. Compared to the chest-strap heart rate monitors, these heart rate monitors are more user friendly and more comfortable to wear. However, studies about their validity are ambiguous [5,6] and focussed on the evaluation of exercise intensity expressed as energy expenditure while neglecting the heart rate component [7,8]. A review by Duking et al. [9] identified only two wrist-worn heart rate monitors that were adequately tested and validated (Polar V800, Mio Alpha 2). They stated that all other heart rate monitors were not scientifically validated and warranted some caution when interpreting the provided information [9]. Moreover, the performance of these wrist worn heart rate monitors depends on various external factors such as: the activity itself, the sensor type and the positioning of the watch [5]. In addition, concurrent with the chest-strap heart rate monitors, the validity of these wrist worn heart rate monitors can be compromised during exercise when signals are more susceptible to motion artefacts. In this regard, Jo

CONTACT Jomme Claes jomme.claes@kuleuven.be Department of Cardiovascular Sciences, KU Leuven, UZ Leuven, Herestraat 49, 3000 Leuven, Belgium

et al. [10] concluded that the two wrist-worn heart rate monitors (Basis Peak, Fitbit Charge HR) they evaluated became less valid or even invalid with increasing exercise intensity.

Since exercise intensity is the most important determinant for health benefits of exercise [11] and heart rate remains the preferred method to prescribe and monitor training intensity [12], these heart rate monitors need to be valid. Therefore, we aimed to evaluate the concurrent criterion validity of the Garmin Forerunner® 225 (Garmin International, Kansas City, MO) for measuring heart rate during walking activities in healthy adults. The Garmin Forerunner® 225 is a commercially available wrist worn heart rate monitor that uses a green light optical sensor, to detect heart rate [13]. Although Garmin plays a prominent role in the world market of wearable devices [14], little is known about the validity of its heart rate monitors.

2. Methods

2.1. Participants

A convenience sample of healthy adults was recruited among colleagues, friends and family of the investigators. Participants had to be regularly physically active men or women between the ages of 20 and 40 years without any known musculoskeletal pathology or cardiovascular, respiratory or metabolic diseases. The study adhered to the guidelines set forth by the declaration of Helsinki and was approved by the UZ Leuven biomedical ethics committee. All participants provided written informed consent.

2.2. Design

In this validity study, participants visited the laboratory in the University Hospital Leuven (UZ Leuven, Belgium) twice. Upon arriving at the laboratory, demographic and anthropometric data, as well as a resting electrocardiogram (ECG) and data on physical activity behaviour (International Physical Activity Questionnaire (IPAQ [15])) were collected. The first visit was a familiarisation session to determine the individual walking speeds that were required to achieve a moderate intensity between 4 and 6 metabolic equivalents (METs) and a high intensity above seven METs [16]. To this purpose, patients were equipped with the Jaeger Oxycon Mobile (Jaeger, Carefusion 234, GmbH, Hoechberg, Germany) for breath by breath gas exchange measurements (ventilation, oxygen consumption, carbondioxide production) and the speed of the treadmill was adjusted until the corresponding oxygen uptake for 4–6 METs (14–21 mlO_2/kg/min) and seven METs or more ($\geq$24.5 mlO_2/kg/min) was reached. The inclination of the treadmill was 5% and 8% in the moderate intensity and high intensity condition respectively and was the same for each participant.

Table 1. Protocol.

Time	Activity
5 minutes	Sitting
5 minutes	Standing
10 minutes	Walking at a speed of 4 km/hour
1 minute	Standing
3 minutes	Sitting
1 minute	Standing
10 minutes	Walking at an inclination of 5% and an intensity between 4 and 6 METs
1 minute	Standing
3 minutes	Sitting
1 minute	Standing
10 minutes	Walking at an inclination of 8% and an intensity above seven METs

During the second visit, participants completed a standardised treadmill protocol. They were asked to perform a series of sitting, standing and walking activities for 50 min as described in Table 1. For analysis, we used the last three minutes of the standing bout and each walking bout to give the heart rate adequate time to stabilise. Average heart rate for this three-minute period obtained with the Garmin Forerunner® 225 was then compared to the average heart rate obtained from an ECG over the same period (Zensor®, Intelesens Ltd, Belfast, UK). The Garmin Forerunner® 225 was programmed with the participants' gender, age, weight and height and was fitted on the left forearm according to the user manual. The FDA approved wireless 3-lead ECG was attached on the chest with the studded attachment electrode placed directly under the left side of the rib cage and the two lead-electrodes placed on both processi coracoideii at the level of the shoulder and served as the criterion method. The electrodes were standard diagnostic quality and had Ag-AgCl, 0.9%NaCl Hydrogel and stud based attachment. Figure 1(A,B) illustrates the Garmin Forerunner® 225 and Zensor®, respectively [17,18]. The Jaeger Oxycon Mobile was used to measure oxygen uptake breath by breath during the protocol to verify and confirm the exercise intensities. The Garmin Forerunner® 225 was started simultaneously with the start of the test. This time point was also manually written down by a second researcher to allow identification of the start point of the test in the Zensor® data. Raw heart rate data were obtained offline through the Zensor + software which provided beat-by-beat information based on the distance

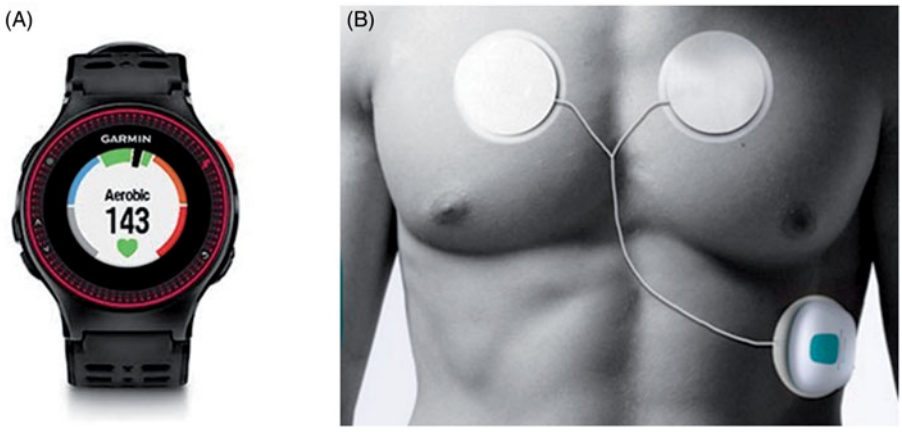

Figure 1. Garmin Forerunner 225 and Zensor. (A) Garmin Forerunner® 225, n.d. image by unknown. (B) Zensor, n.d. image by unknown.

between two R-waves on the ECG. Data from the Garmin Forerunner® 225 were downloaded to the web-based manufacturer application, i.e. the Garmin Trainingcenter, which allowed access to heart rate averages over a randomly ranging amount of seconds, independent of the exercise intensity.

2.3. Statistical analysis

SPSS (version 20; SPSS for windows; SPSS Inc., Chicago, IL) was used to analyse the data and Graphpad Prism 5.0 (Graphpad Software, San Diego, CA) to plot the figures. Descriptive data are reported as mean and standard deviation (SD) or as median and range. A paired *t*-test was used to compare the Garmin and ECG derived mean heart rates for each bout. Concurrent validity was examined by calculating Pearson's correlation coefficients (*r*). A *r*-value of 0–0.29 was considered as no correlation, 0.30–0.49 as a weak correlation, 0.50–0.69 as moderate and 0.70–1.00 as a strong correlation. Pearson's correlation was also performed on averaged heart rate values per minute during the complete protocol. A Bland–Altman plot was constructed including the results of each condition with 95% limits of agreement (LoA) to allow the assessment of the accuracy of the Garmin Forerunner® 225. Root mean squared error (RMSE) was calculated and expressed both as beats per minute (bpm) and percentage to assess the accuracy of the Garmin Forerunner® 225 in comparison with the ECG criterion measurements [19,20]. The formulas used can be found below. Significance was set at an alpha level of $p < .05$ two-tailed.

$$RMSE\ (bpm) = \sqrt{\frac{\sum_{k=1}^{n} (HRgarmin\ [k] - HRzensor\ [k])^2}{n}}$$

$$RMSE\ (\%) = \sqrt{\frac{\sum_{k=1}^{n} \left(\frac{(HRgarmin\ [k] - HRzensor\ [k])}{HRzensor[k]} * 100\%\right)^2}{n}}$$

3. Results

In total, a convenience sample of 12 physically active volunteers (six men, six women) was included in the study. The mean age was 28 years (SD: 4.79). Nine participants were Caucasian and three were South-Americans (one Suriname, one Colombia and one Brazil). Participant characteristics are shown in Table 2.

Table 3 shows the mean heart rate during the last three minutes of rest and each walking bout. We observed no statistically significant difference between the mean heart rates of the Garmin Forerunner® 225 and the criterion method ($p > .05$ for all). Further, as can be derived from Table 4, Pearson's correlations between both devices were moderate to strong for each condition and for the complete test.

The Bland–Altman plot of heart rates is illustrated in Figure 2 and shows a general underestimation of the heart rate by Garmin. The bias is −1.57 bpm and LoA range from −32.53 to 29.40 bpm. In Figure 3, which shows the average heart rates per minute for both devices, an underestimation of the heart rate by the Garmin could be seen especially at higher intensities. However, a slight overestimation by the Garmin was present at the lower end of the heart rate spectrum. The RMSE showed a deviation of 3.01 bpm (2.89%).

4. Discussion

The results of the correlation analysis, paired *t*-tests and the Bland–Altman plot suggested good

Table 2. Basic characteristics of participants.

	Mean	Standard deviation
Age (years)	28	4.79
Weight (kg)	69.28	13.24
Height (cm)	176.23	10.97
BMI (kg/m^2)	22.14	3.46
Average walking speed 5% inclination (km/h)	4.82	0.29
Average walking speed 8% inclination (km/h)	5.68	0.20
IPAQ-category [median (range)]	2	1–3

BMI: body mass index; IPAQ: International Physical Activity Questionnaire. All values are reported as means ± standard deviations unless stated otherwise.

Table 3. Comparison of mean heart rates between Garmin and ECG at different time points.

Condition	Garmin (bpm)	Zensor (bpm)	*p* Value
Rest	90 ± 15	89 ± 21	.881
Walking 4 km/h	101 ± 24	99 ± 24	.718
Walking 5% inclination	116 ± 25	113 ± 23	.456
Walking 8% inclination	141 ± 27	143 ± 28	.767

Data are reported as mean ± SD.

Table 4. Pearson's correlation for each of the four conditions.

Time	*r*	*p* Value (two-tailed)
At rest	0.822	.001
Walking 4 km/h	0.650	.022
Walking 5% inclination	0.868	<.001
Walking 8% inclination	0.769	.003
Complete protocol	0.947	<.001

r=Pearson's correlation coefficient.

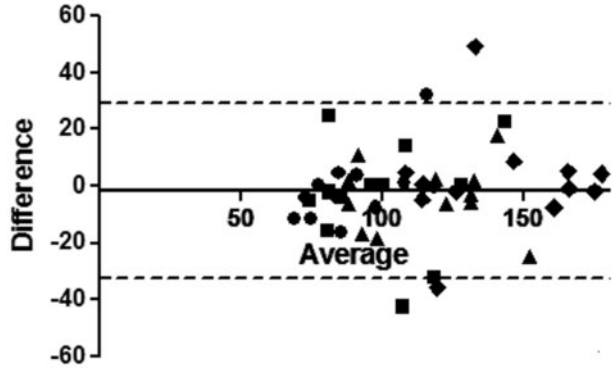

Figure 2. The Bland–Altman plot of heart rate measurements of Garmin vs Zensor. The full horizontal line is the bias, the dotted lines are the limits of agreement. Resting measurements are represented by dots, low intensity walking by squares, moderate intensity walking by triangles and high intensity walking by diamonds.

agreement between the Garmin Forerunner® 225 and ECG, however, the wide LoA warrant some caution when using the Garmin Forerunner® 225 in a setting that requires very precise heart rate measurements. In the current literature, conclusions about validity are often made based on correlation coefficients and

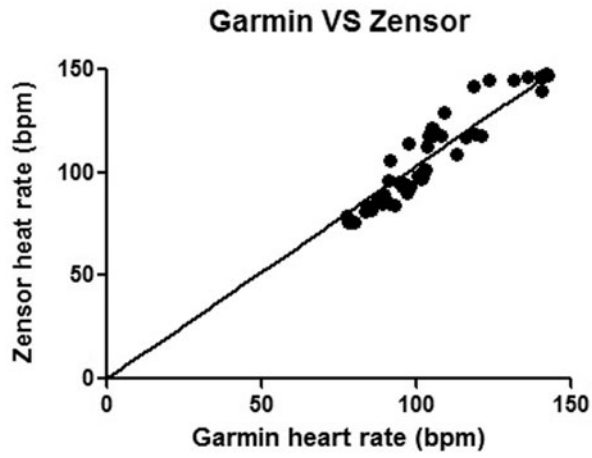

Figure 3. Average heart rates per minute for the complete test of each participant are depicted for both devices/bpm: beats per minute.

mean biases of the Bland–Altman plots. LoA provided by the Bland–Altman statistics are discussed, but rarely affect the final conclusion [6,10,21,22]. This can result in identifying a heart rate monitor as valid, when simultaneously large inter-individual differences in heart rate are observed. However, when assessing an individual person, the capability of a heart rate monitor to measure heart rate correctly with a small error margin is more important.

In line with our work, Wallen et al. [21] reported a non-significant difference between heart rate measurements in a study evaluating the accuracy of four PPG-based watches during rest, walking/running and cycling (ranging from −1 to −9%) [21]. Intraclass correlation values ranged from 0.78 to 0.98 suggesting excellent agreement between each of the four watches and the ECG criterion method. Although their LoA ranged between −27.3 bpm and 13.1 bpm and were somewhat smaller than found here, they highlight the fact that some caution is needed when using these watches to individually and accurately guide athletes or patients during training. Parak and Korhonen [5] also showed an underestimation of the heart rate measured by the MIO Alpha during varying activities. A mean error ranging from −0.52% to −4.80% was reported with the largest error measured during cycling [5]. In the same study, the Scosche myRythm was tested and showed mean errors ranging from −3.13% to 0.63% with the largest error during walking [5].

The variability in measurements of PPG based heart rate monitors could be due to the observation that the quality of PPG data is sensitive to external factors such as large movements [23]. Such movements were

present in our research because the Garmin Forerunner® 225 was worn around the left wrist and participants heavily used their arms while walking on an inclination of 5% and 8%. This helps explain why during high intensity walking more outliers were present in the Bland–Altman plot compared to the other conditions of the protocol. However, other activities may require even larger or faster movements of the arms which could hinder adequate heart rate measurement. Therefore, it could be hypothesised that during running a less accurate heart rate measurement will be acquired. Despite the name of the Garmin Forerunner® 225, we did not test this heart rate monitor during running activities. The reason for this was twofold: (1) the participants wore two modules of the Jaeger Oxycon Mobile on the chest which greatly hindered running comfortably and (2) to limit the amount of motion artefact in the ECG recordings. Another external factor influencing PPG sensitivity is sweat [24]. Participants performed the protocol indoors in a room with a temperature between 19 °C and 23 °C and this resulted in intense perspiration in some of them, possibly interfering with adequate heart rate measurements by the Garmin heart rate monitor. However, these are parameters that can also affect the measurement accuracy of normal ECG electrodes [25]. Furthermore, underlying pathologies could possibly interfere with the accuracy of heart rate registration. Cardiac patients with frequent extrasystoles or arrhythmias could therefore easily misinterpret their exertion level during exercise since the detection of arrhythmias is not yet incorporated in these watches. However, recent studies have shown that arrhythmia monitoring with the use of PPG has become possible [26].

The absence of these external factors could explain why our results indicated that the *r* between both methods was highest in rest and for exercise at intensities between 4 and 6 METs. The wide dispersion of heart rates in the Bland–Altman plot during the low intensity phase was unexpected and we are currently unable to explain this. However, a remark has to be made regarding the relative aspect of the intensity levels. Our participant group consisted of people with varying fitness and activity levels. As the speed that was needed to walk at 4–6 METs against a 5% inclination was measured by means of oxygen consumption (14–21 mlO_2/min/kg), this might have been moderate intense for a part of our population, but for some of them this was probably not even 40% of their maximum oxygen uptake. Therefore, some of our values that were taken into account to calculate the average moderate intensity heart rate were actually low intensity values. Furthermore, the results of this study do not suggest that all watches using PPG-technology to assess heart rate will provide accurate heart rate measurements. That is, other PPG-based heart rate monitors may use different optical sensors, algorithms and/or infra-red wavelengths [13,27]. Our study was conducted in a small study population involving only three individuals with a somewhat darker skin complexion. It is known that darker skin tones can disturb the PPG signal and we recommend more research into the validity of these wrist worn heart rate watches in individuals of different ethnicities [13]. Another limitation of the small study population is that one outlier could potentially influence results. We saw one consistent outlier when we constructed the Bland–Altman plots but removal of this individual did not change the correlation values. Upon analysis of the heart rate values at the four different time points of this participant it could be concluded that the registration by the Garmin Forerunner® 225 reported unexpected values. This participant was a fit girl with a BMI of 18.3 kg/m^2 and very thin wrists. We suspected that the Garmin Forerunner® 225 could not be attached firmly enough, resulting in invalid heart rate values which is a common problem due to optical coupling with the skin. Finally, given our large LoA, some caution might be warranted from a clinical perspective, i.e. if one aims to rely on these heart rate monitors to guide patients while training, some patients might be overreaching their maximal target heart rate. These rather large LoA might also be due to our small sample size. Nevertheless, even in studies with larger sample sizes, LoA tend to be wide depending on the exercise intensity and on the heart rate monitor tested [10,22,28]. To this end, more research is needed in a large sample size to confirm our result.

Disclosure statement

The authors report no conflict of interest.

Funding

This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation Action under Grant Agreement no. 643491. PATHway: Technology enabled behavioural change as a pathway towards better self-management of CVD (www.pathway2health.eu). The research was conducted at the University Hospital Leuven (UZ Leuven), Herestraat 49, 3000 Leuven.

ORCID

Jomme Claes http://orcid.org/0000-0003-3922-0038

References

[1] Thompson WR. Worldwide survey of fitness trends for 2016: 10th anniversary edition. ACSM'S Health Fitness J. 2015;19:9–18.

[2] Garber CE, Blissmer B, Deschenes MR, et al. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc. 2011;43:1334–1359.

[3] Weippert M, Kumar M, Kreuzfeld S, et al. Comparison of three mobile devices for measuring R-R intervals and heart rate variability: polar S810i, Suunto t6 and an ambulatory ECG system. Eur J Appl Physiol. 2010;109:779–786.

[4] Kingsley M, Lewis MJ, Marson RE. Comparison of Polar 810s and an ambulatory ECG system for RR interval measurement during progressive exercise. Int J Sports Med. 2005;26:39–44.

[5] Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photoplethysmography. Conf Proc IEEE Eng Med Biol Soc. 2014;2014: 3670–3673.

[6] 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.

[7] Bai Y, Welk GJ, Nam YH, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc. 2016;48:151–158.

[8] Lloyd LK, Crixell SH, Price LR. Is the Polar F6 heart rate monitor less accurate during aerobic bench stepping because of arm movements? J Strength Cond Res. 2014;28:1952–1958.

[9] Duking P, Hotho A, Holmberg HC, et al. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. Front Physiol. 2016;7:71.

[10] Jo E, Lewis K, Directo D, et al. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med. 2016;15:540–547.

[11] Swain DP, Franklin BA. Comparison of cardioprotective benefits of vigorous versus moderate intensity aerobic exercise. Am J Cardiol. 2006;97:141–147.

[12] Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr. 2014;1:5.

[13] Fallow BA, Tarumi T, Tanaka H. Influence of skin type and wavelength on light wave reflectance. J Clin Monit Comput. 2013;27:313–317.

[14] Shirer MLR. Worldwide wearables market increases 67.2% amid seasonal retrenchment, according to IDC Framingham: IDC; 2016 [cited 2017 Jan 25]. Available from: http://www.idc.com/getdoc.jsp?containerId=prUS41284516

[15] Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35:1381–1395.

[16] Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc. 2011;43:1575–1581.

[17] Garmin Forerunner 225, n.d. photograph, viewed 9 Nov 2016. Available from: https://buy.garmin.com/nl-NL/NL/sport-recreatie/hardlopen-fitness/vorige-modellen/forerunner-225/prod512478.html

[18] Zensor, n.d. photograph, viewed 9 Nov 2016. Available from: http://www.intelesens.com/archivedSite/inhomemonitoring/zensor.html

[19] Hyndman RJKA. Another look at measures of forecast accuracy. Int J Forecast. 2006;22:679–688.

[20] Poh MZ, Poh YC. Validation of a standalone smartphone application for measuring heart rate using imaging photoplethysmography. Telemed J E-Health. 2017;23. doi: 10.1089/tmj.2016.0230.

[21] Wallen MP, Gomersall SR, Keating SE, et al. Accuracy of heart rate watches: implications for weight management. PLoS ONE. 2016;11:e0154420.

[22] 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. doi: 10.1136/bmjsem-2015-000106

[23] Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. J Med Syst. 2011;35: 969–976.

[24] Maeda Y, Sekine M, Tamura T. The advantages of wearable green reflected photoplethysmography. J Med Syst. 2011;35:829–834.

[25] Crawford JDL. Practical aspects of ECG recording. Keswick: M&K Update Ltd; 2012.

[26] Carpenter A, Frontera A. Smart-watches: a potential challenger to the implantable loop recorder? Europace. 2016;18:791–793.

[27] Zhang Z, Pi Z, Liu B. TROIKA: a general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. IEEE Trans Biomed Eng. 2015;62:522–531.

[28] Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol. 2017;2: 104–106.