Comparison Between Smartwatch-Derived and CPET-Measured VO_2max

Alexandra Jamieson[1], Siana Jones[1], Claire Steves[2], Nicholas Timpson[3], Nishi Chaturvedi[1], Alun D Hughes[1], Michele Orini[1,4]

[1]MRC Unit for Lifelong Health and Ageing, UCL, UK
[2] Department of Twin Research and Genetic Epidemiology, King's College London, UK
[3] MRC Integrative Epidemiology Unit, University of Bristol, UK
[4] Department of Biomedical Engineering, King's College London, UK

Abstract

This study aimed to determine the accuracy of smartwatch-derived maximal oxygen consumption (VO_2max) and the percentage of predicted VO_2max (%pVO_2max), two standard measures of cardiorespiratory fitness with established clinical predictive value. 215 adults (44 (21%) male; median [interquartile range; IQR] 55 [32, 62] years old) performed a maximal exercise test (CPET) on a semi-recumbent ergometer and wore a Garmin Vivoactive 4s (GV4) smartwatch for 60 days. The first and last VO_2max estimates provided by GV4 were compared to CPET measured VO_2max and %pVO_2max (Wasserman and Whipp's anthropometric-based equation). Agreement was assessed using Bland-Altman analysis (bias and limits of agreement [LoA]), absolute percentage error (APE), reported as median [interquartile range], and Pearson's correlation coefficient (cc). VO_2max and %pVO_2max measured during CPET was 22.4 [17.5, 27.4] ml/kg/min and 90.9% [78.1%, 101.3%], respectively. VO_2max estimates from GV4 were moderately correlated with CPET measures (cc≤0.66) and showed a large positive bias ~14 ml/kg/min with LoA from 0 – 27 ml/kg/min. Correlation between VO_2max from GV4 and anthropometric-based prediction of VO_2max was high (cc>0.90). Agreement between %pVO_2max from GV4 and CPET was poor (cc~0.15, bias ~52%, LoA 7-98 %). GV4 provides estimates of VO_2max that overestimate but moderately correlate with CPET measured VO_2max. The agreement for %pVO_2max is poor.

1. Introduction

Cardiorespiratory fitness (CRF) has been linked to several health-related outcomes, with low fitness being associated with increased risk of cardiovascular disease [1, 2], metabolic syndrome [3], cognitive function [4] and severe COVID-19 [5].

Maximal cardiopulmonary exercise testing (CPET) is considered the gold standard assessment of CRF. It is a dynamic symptom-limited test that is incremental in nature with a continual increase in workload until the individual cannot exercise anymore (self-reported exhaustion) [6]. The CPET allows for breath-by-breath analysis of gas exchange: oxygen consumption (VO_2) and carbon dioxide production (VCO_2) to derive the primary outcome measure of maximal oxygen consumption (VO_2max). Direct measurement of VO_2max requires expensive monitoring equipment, experienced personnel and is not without risk. Therefore, the application of maximal CPET is limited in the context of the unselected general population or large epidemiological studies. Furthermore, high levels of motivation and physical effort are required by the individual to achieve a maximal test, which may not be feasible in the presence of chronic conditions such as pain or fatigue.

Predicted VO_2max is typically estimated using anthropometric-based equations which are population specific, based on both active and sedentary individuals, men and women and individuals with and without cardiac conditions [7, 8]. The percentage of predicted VO_2max (%pVO_2max) is calculated as the ratio between measured or estimated VO_2max and predicted VO_2max. %pVO_2max has both prognostic value in assessing CRF and predicting clinical outcomes [9].

Novel wrist-worn wearable technologies (hereafter *smartwatches*) use tri-axial accelerometers, and photoplethysmography (PPG) sensors to measure physiological parameters such as heart rate (HR), distance and step count. Smartwatch estimates of VO_2max are derived using the PPG signal measured HR while adjusting for factors such as age, sex, height, weight and exercise type using proprietary algorithms [10, 11]. In the context of healthcare, these devices provide an opportunity to estimate CRF parameters outside of the clinical environment, at scale.

Molina-Garcia and colleagues (2022) performed a systematic review with meta-analysis of 14 studies (n=403) that assessed the validity of smartwatch estimation of VO_2max in both resting and exercise test conditions [12]. In the context of resting conditions, the authors observed an overestimation of VO_2max (Bias [Limits of

Agreement; LoA]= 2.17 [-13.07, 17.41] ml/kg/min; p=0.020) compared to the reference CPET. In contrast, a bias close to nil but wide LoA (Bias [LoA]= -0.09 [-16.79, 16.61] ml/kg/min; p=0.910) was observed when exercise test conditions were utilized.

Yet, there are very few studies assessing the agreement between CPET (reference standard) assessment of CRF and smartwatch estimation of VO_2max from remote community-based data capture. The aim of this study was to determine smartwatch device accuracy in estimating VO_2max and %pVO_2max using data from CPET as a reference.

2. Methods

2.1. Study participants

215 adults (44 (21%) male; median [interquartile range; IQR] 56 [32, 62] years old) were recruited from two population-based cohorts the Avon Longitudinal Study of Parents and Children (ALSPAC; REC: 21/SC/0030) [13, 14] and TwinsUK (REC: 19/NW/0187). Clinic investigations were conducted at the UCL Bloomsbury Centre for Clinical Phenotyping, London. All participants gave written informed consent.

Participant age and sex were collected by questionnaire. Height was measured using a stadiometer (Seca217, Seca, Germany) to the closest centimetre and weight was measured in kilograms using digital bio-impedance scales (BC-418 or MC-780MA, Tanita, USA) to calculate BMI.

2.2. Cardiopulmonary exercise test

All exercise tests were conducted according to the ATS/ACCP (2003) guidelines for CPET [15]. CPET was performed using a semi-recumbent cycle ergometer (Ergoline 900, Hamburg, Germany) and metabolic cart (Quark Cosmed, Rome, Italy). Following 1 minute of rest and 2 minutes of warm-up cycling at 5 Watts, the work rate was incrementally increased using a ramp protocol in an individualized manner by either 15, 20, 25 or 30 Watts each minute based on the Wasserman weight algorithm [16]. The test was terminated if the participant (i) reached their age-predicted (220-age) maximum HR (ii) experienced limiting symptoms or (iii) developed arrhythmia, hypotension (systolic blood pressure (BP) drop of >10mmHg despite increasing workload) or (iv) an excessive blood pressure rise during the test (>250 systolic mmHg). Expired gases were analyzed breath-by-breath, and HR measured using a continuous 6-lead ECG (Quark CPET, Cosmed, Italy). Key CPET outcome measures derived included highest achieved VO_2 (VO_2max) and peak HR. %pVO_2max was estimated as the ratio between measured or estimated VO_2max and predicted VO_2max using sex-specific equations from Wasserman and Whipp [8]. 145 participants wore the GV4 during CPET.

2.3. Smartwatch VO_2max estimate

Participants were fitted with a Garmin Vivoactive 4s (GV4) smartwatch. An app enabled the transfer of physiological data from the watch to servers at UCL for research purposes [17]. Garmin reports that VO_2max estimates can be generated from all-day passive HR data collection, activity recordings or both and can take up to 30 days to populate.

2.4. Data processing & statistical analysis

First and last VO_2max estimates provided by GV4 along with the number and duration of activity recordings in the 60 days following CPET were used. %pVO_2max was measured using predicted VO_2max from Wasserman and Whipp's anthropometric-based equations [16]. Statistical analyses were performed using MATLAB 2022a. Sample characteristics and outcome parameters are described using median [interquartile range; IQR] for continuous variables and frequency (percentage) for categorical variables. Agreement was assessed using Bland-Altman analysis [18] presented as mean bias [Limits of agreement; LoA], absolute percentage error (APE), reported as median [IQR] and Pearson's correlation coefficient (cc). LoAs were measured as $m_E \pm 1.96*SD_E$, where m_E and SD_E represent the mean and standard deviation of the estimation error.

3. Results

A summary of study participant characteristics and study parameters is presented in Table 1. The peak HR and peak VO_2 measured during CPET were 145 [129, 159] beats per minute (bpm) and 22.4 [17.5, 27.4] ml/kg/min, respectively.

Table 1. Study participant characteristics and outcome parameters presented as median [IQR; interquartile range] or n (%). BMI (body mass index), HR (heart rate), bpm (beats per minute), CPET (cardiopulmonary exercise test), GV4 (Garmin Vivoactive 4s), VO_2 (oxygen consumption), No. (number).

	n	Median [IQR] or n (%)
Age (years)	215	56 [32, 62]
Male Sex	215	44 (21%)
BMI (kg/m^2)	215	25.1 [21.9, 28.5]
Resting HR (bpm)	212	70 [62, 78]
CPET peak HR (bpm)	212	145 [129, 159]
GV4 peak HR (bpm)	145	148 [134, 163]
CPET VO_2max (ml/kg/min)	207	22.4 [17.5, 27.4]
GV4 VO_2max [First] (ml/kg/min)	215	35.0 [30.0, 41.9]
GV4 VO_2max [Last] (ml/kg/min)	215	37.0 [31.3, 43.0]

CPET %pVO₂max	144	91 [78, 101]
GV4 %pVO₂max [First]	207	142 [135, 149]
GV4 %pVO₂max [Last]	207	148 [138, 155]
No. GV4 activities	214	18 [10, 42]
No. GV4 activities ≥10 mins	214	9 [2, 32]
Activity duration (mins)	214	21 [9, 35]

VO_2max estimates from GV4 were moderately correlated with CPET measures (cc=0.62 and 0.66 for first and last estimates) and showed a large positive bias ~14 ml/kg/min with LoA from 0 – 27 ml/kg/min. Correlation between VO_2max from GV4 and anthropometric-based prediction of VO_2max was high (cc>0.90). Agreement between %pVO_2max from GV4 and CPET was poor (cc~0.15, bias ~52%, LoA 7-98 %) (Table 2 & Figure 1).

Table 2. Level of agreement between Garmin Vivoactive 4s (GV4) derived and cardiopulmonary exercise test (CPET) measured parameters presented as Pearson's correlation coefficients (cc), bias [LoA; Limits of Agreement] and absolute percentage error (APE) median [interquartile range; IQR]. VO_2 (oxygen consumption), Ave. (average), %p (% predicted), HR (heart rate), bpm (beats per minute).

	n	cc	Bias [LoA]	APE [IQR]
VO_2max [First] (ml/kg/min)	207	0.62	13.4 [0.3, 26.5]	56 [40, 80]
VO_2max [Last] (ml/kg/min)	207	0.66	14.6 [2.5, 26.6]	60 [45, 88]
%pVO_2max [First]	207	0.14	52.4 [7.0, 97.8]	56 [40, 80]
%pVO_2max [Last]	207	0.28	57.4 [15.5, 99.2]	60 [45, 88]
Peak HR (bpm)	144	0.93	3.9 [-11.9, 19.7]	4 [2, 7]

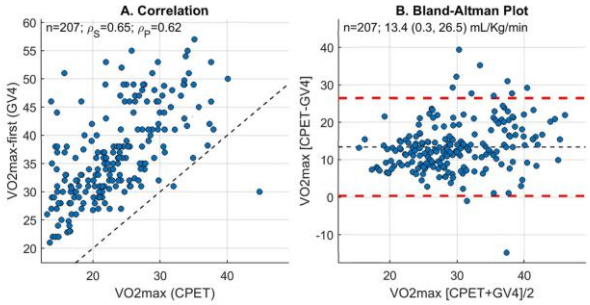

Figure 1. Correlation and Bland-Altman plots demonstrating levels of agreement between Garmin Vivoactive 4s (GV4) derived VO_2max [First] and cardiopulmonary exercise test (CPET) measured VO_2max.

4. Discussion

We sought to establish the accuracy of remote smartwatch estimation of VO_2max and %pVO_2max compared to a clinic based CPET. Our primary findings were 1) a moderate correlation and agreement with large positive bias between GV4 derived and CPET-measured VO_2max; 2) no improvement in GV4 performance after two months of monitoring; 3) agreement between GV4 derived and CPET measured %pVO_2max was poor.

Clinical assessment of CRF provides an optimal approach for stratifying patients according to risk [9] and smartwatches provide an opportunity to do so remotely without the requirement for expensive testing equipment, clinical staff and time. We observed moderate agreement with large positive bias between GV4 derived and CPET measured VO_2max, consistent with Molina-Garcia and colleagues' review findings during resting conditions [12]. HR is the primary parameter utilised by smartwatches for the estimation of VO_2max. In line with prior work, we observed strong correlation, good agreement and small APE between GV4 derived and CPET measured HR [19]. A limitation of our study is that CPET was performed on a semi-recumbent cycle ergometer which can be associated with a lower achieved VO_2max when compared to that of treadmill testing [20] or upright cycling [21]. We cannot confirm whether the observed differences between the GV4 derived and CPET measured VO_2max are due, in part, to testing modality.

We observed poor agreement between GV4 derived and CPET measured %pVO_2max. This may in part be explained by the strong correlation (cc~0.90) between VO_2max estimated by GV4 and by anthropometric-based equations. As expected, CPET-measured VO_2max was lower, but GV4-estimated VO_2max was higher, than anthropometric-based prediction of VO_2max. This in turn, resulted in poor agreement between CPET- and GV4-based %pVO_2max.

Other limitations of our study include that all participants wore the GV4 and therefore our results may not be generalizable to other manufacturers. Participants were not asked to perform strenuous physical activities whilst wearing the GV4, which may have impacted the accuracy of VO_2max estimation. A comparison between those who did and did not wear the GV4 during CPET was not performed and would be valuable.

5. Conclusion

GV4 provided estimates of VO_2max that overestimates but moderately correlated with CPET measured VO_2max. The agreement for %pVO_2max was poor.

Acknowledgments

The UK MRC and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. TwinsUK is funded by the Wellcome Trust, MRC, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation

(CDRF), Zoe Ltd, the NIHR Clinical Research Network and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

References

1. Blair, S.N., et al., Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. Jama, 1996. **276**(3): p. 205-210.
2. Kodama, S., et al., Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. Jama, 2009. **301**(19): p. 2024-2035.
3. Zaccardi, F., et al., Cardiorespiratory fitness and risk of type 2 diabetes mellitus: A 23-year cohort study and a meta-analysis of prospective studies. Atherosclerosis, 2015. **243**(1): p. 131-137.
4. Themanson, J. and C. Hillman, Cardiorespiratory fitness and acute aerobic exercise effects on neuroelectric and behavioral measures of action monitoring. Neuroscience, 2006. **141**(2): p. 757-767.
5. Ekblom-Bak, E., et al., Cardiorespiratory fitness and lifestyle on severe COVID-19 risk in 279,455 adults: a case control study. International Journal of Behavioral Nutrition and Physical Activity, 2021. **18**: p. 1-16.
6. McArdle WDK, F.I.K., Victor L Exercise physiology : nutrition, energy and human performance. 7th revised International ed. 2009, Philadelphia: Lippincott Williams and Wilkins.
7. Bruce, R.A., F. Kusumi, and D. Hosmer, Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. American heart journal, 1973. **85**(4): p. 546-562.
8. Wasserman, K., et al., Principles of exercise testing and interpretation: including pathophysiology and clinical applications. Medicine and Science in Sports and Exercise, 2005. **37**(7): p. 1249.
9. Ross, R., et al., Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation, 2016. **134**(24): p. e653-e699.
10. Ltd., F.T., Automated Fitness Level (VO2max) Estimation with Heart Rate and Speed Data. 2014.
11. Apple. Using Apple Watch to Estimate Cardio Fitness with VO2 max. 2021 [cited 2024; Available from: https://www.apple.com/healthcare/docs/site/Using_Apple_Watch_to_Estimate_Cardio_Fitness_with_VO2_max.pdf.
12. Molina-Garcia, P., et al., Validity of estimating the maximal oxygen consumption by consumer wearables: a systematic review with meta-analysis and expert statement of the INTERLIVE network. Sports Medicine, 2022. **52**(7): p. 1577-1597.
13. Boyd, A., et al., Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology, 2013. **42**(1): p. 111-127.
14. Fraser, A., et al., Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International journal of epidemiology, 2013. **42**(1): p. 97-110.
15. American College of Sports Medicine, et al., ACSM's guidelines for exercise testing and prescription. Tenth edition. ed. 2018, Philadelphia: Wolters Kluwer. 472.
16. Wasserman K, S.W., Sietsema KE, Sun XG, Whilpp BJ, Principles of Exercise Testing and Interpretation: Including Pathophysiology and Clinical Applications. 2012, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
17. Ranjan, Y., et al., RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR mHealth and uHealth, 2019. **7**(8): p. e11734.
18. Linden, A., RMLOA: Stata module to compute limits of agreement for data with repeated measures. 2021.
19. Fuller, D., et al., Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. JMIR Mhealth Uhealth, 2020. **8**(9): p. e18694.
20. Carter, H., et al., Oxygen uptake kinetics in treadmill running and cycle ergometry: a comparison. Journal of applied physiology, 2000. **89**(3): p. 899-907.
21. Wehrle, A., et al., Power Output and Efficiency During Supine, Recumbent, and Upright Cycle Ergometry. Frontiers in Sports and Active Living, 2021. **3**: p. 161.

Address for correspondence:

Alexandra Jamieson
1-19 Torrington Place, London, WC1E 6HB
alexandra.jamieson@ucl.ac.uk (cc: m.orini@kcl.ac.uk)