# Interchangeability of research and commercial wearable device data for assessing associations with cardiometabolic risk markers

Kingsnorth, A., Moltchanova, E., Thomas, J., Whelan, M., Orme, M. W., Esliger, D. W. & Hobbs, M



**Original citation & hyperlink:**
Kingsnorth, A, Moltchanova, E, Thomas, J, Whelan, M, Orme, MW, Esliger, DW & Hobbs, M 2023, 'Interchangeability of research and commercial wearable device data for assessing associations with cardiometabolic risk markers', Journal for the Measurement of Physical Behaviour, vol. (In-Press), pp. (In-Press).
https://doi.org/10.1123/jmpb.2022-0050

DOI 10.1123/jmpb.2022-0050
ISSN 2575-6605
ESSN 2575-6613

Publisher: Human Kinetics

# Interchangeability of research and commercial wearable device data for assessing associations with cardiometabolic risk markers

**Submission date:** 23/04/2023

# Abstract

**Introduction:** Whilst there is evidence on agreement, it is unknown whether commercial wearables can be used as surrogates for research grade devices when investigating links with markers of cardiometabolic risk. Therefore, the aim of this study was to investigate if data from a commercial wearable device could be used to assess associations between behaviour and cardiometabolic risk markers, compared to physical activity from a research grade monitor.

**Methods:** Forty-five adults concurrently wore a wrist-worn Fitbit Charge 2 and a waist-worn ActiGraph wGT3X-BT during waking hours over 7 consecutive days. Log-linear regression models were fitted, and predictive fit via a 1-out cross-validation was performed for each device between behavioural (steps, light and moderate-to-vigorous physical activity) and cardiometabolic variables (body mass index [BMI], weight, body fat %, systolic and diastolic blood pressure, glycated haemoglobin, grip strength, estimated maximal oxygen uptake and waist circumference).

**Results:** Overall, step count was the most consistent predictor of cardiometabolic risk factors, with negative associations across both Fitbit and ActiGraph devices for BMI (-0.017 vs. -0.020, $p<0.01$), weight (-0.014 vs. -0.017, $p<0.05$), body fat % (-0.021 vs. -0.022, $p<0.01$) and waist circumference (-0.013 vs. -0.015, $p<0.01$). Neither device was found to provide a consistently better prediction across all included cardiometabolic risk markers.

**Conclusions:** Step count data from a commercial grade wearable device showed similar associations and predictive relationships with cardiometabolic risk markers compared to a research-grade wearable device, providing preliminary support for their use in health research.

**Words:** 239



**Keywords:** Commercial wearables, step count, cardiometabolic risk markers

# 1. Introduction

In behavioural measurement, a diverse range of methods exist that estimate physical behaviours ranging from self-reported questionnaires to wearable devices (Esliger et al., 2017). Wearables can further be split in two broad categories of research-grade and commercial devices, yet the cost and feasibility of deploying research-grade wearables often prevents them from being used within large studies. With nearly two in five (38%) individuals aged 35-54 in the United Kingdom reporting owning a smart watch or wearable fitness tracker (OFCOM, 2021), an opportunity exists for capturing organic data about physical behaviours over multiple weeks or months (Pontin et al., 2021). This is an exciting proposition given the potential for population surveillance and health status risk profiling prediction on a large scale (Strain et al., 2019).

Epidemiological evidence has demonstrated that being physically active is associated with reduced risks of developing several diseases and all-cause mortality, and generally having a more favourable cardiometabolic risk factor profile (Chastin et al., 2021; Hajna et al., 2018; Huang et al., 2020; Janssen et al., 2020; Warburton et al., 2006). However, there are few studies investigating the associations between physical activity and cardiometabolic variables using data captured by wearables, rather than data from research grade devices or questionnaires. Using step and intensity-based metrics from a wearable device, positive relationships have been shown with high-density lipoprotein (HDL), body mass index (BMI) and waist circumference (Rykov et al., 2020). Having a higher resting heart rate and a greater sleep efficiency has also been associated with a lower BMI and waist circumference (Lim et al., 2018; Teo et al., 2019). Despite the potential link between wearable metrics and cardiometabolic variables, the results are not compared to a ground truth measure (a research-grade device).

If models derived from commercial devices have similar model parameters to those derived from research grade devices, wearable data has the potential in being used more commonly in larger studies concentrating on health and physical behaviours. The aim of this study was therefore to investigate if data from a commercial wearable device could be used to assess associations between behaviour and cardiometabolic risk markers, compared to physical activity derived from a research grade monitor.

# 1. Methods

## 2.1 Study setting and ethics

This study was a secondary data analyses using data that were obtained as part of the Sensing Interstitial Glucose to Nudge Active Lifestyle (SIGNAL) interventional trial (Whelan et al., 2017, 2021). Data were collected between July and October 2017 and ethical approval was provided from Loughborough University Ethics Advisory Committee (R17-P049).

## 2.2 Participants and procedure

Participants were 45 adults (60% female) aged 40 or older, recruited from the community in Leicestershire and classified with a moderate-to-high risk of developing type 2 diabetes using the Leicester Risk Assessment Tool (Gray et al., 2010). Participants also had a compatible Android smartphone to facilitate participation due to restrictions in glucose monitoring compatibility at the time. Written informed consent was obtained from all participants prior to taking part. Participants wore both a research-grade device and commercial wearable device during the baseline phase of the study.

## 2.3 Measurements

2.3.1 Physical behaviours

Participants were requested to wear a wrist-worn Fitbit Charge 2 (Fitbit, San Francisco, USA) and a waist-worn ActiGraph wGT3X-BT (ActiGraph, Pensacola, USA) during waking hours over 7 consecutive days.

The Fitbit was deployed during an in-person appointment on the non-dominant wrist and participants were provided with study account credentials to sync their data. This enabled the research team to access participant data and check syncing adherence via Fitabase (Small Steps Labs LLC, San Diego, USA). Several day level movement behaviour variables from the Fitbit were obtained: step count and minutes spent within sedentary, lightly active (<3 metabolic equivalents [METs]), fairly active (3-5.9 METs) and very active intensities (≥6 METs) (Van Blarigan et al., 2017). For this study, the fairly active and very active intensities were collapsed to derive the more commonly used, and physiologically relevant, moderate to vigorous physical activity (MVPA) intensity category. A step threshold per day of ≥1500 steps/day was used as a valid day criteria in line with previous studies (Chu et al., 2017; Kingsnorth et al., 2021; Mikkelsen et al., 2020; Tudor-Locke et al., 2015). Day level summaries were exported for each participant. In line with other studies that have used wearable data from Fitbit devices, sedentary time was not included due to contamination by non-wear time (Kingsnorth et al., 2021).

Using an elasticated, nylon waist belt, the ActiGraph device was deployed over the right hip (midclavicular line), and devices were set to record data at 100 Hz using ActiLife (ActiGraph, Pensacola, USA). Data files were integrated into 60 second epoch .agd files and processed using KineSoft (Kinesoft v3.3.80, Loughborough, UK). A valid day was defined as 600 minutes of count data

per day and non-wear was defined as 60 minutes of consecutive zero values, with allowance of up to 2 minutes of interruptions (Troiano et al., 2008). The intensity of physical activity was defined with the following cut points: light activity (100-2019 cpm) and MVPA (≥2020 cpm) (Troiano et al., 2008).

2.3.2 Cardiometabolic risk markers

During the initial visit of the SIGNAL study, commonly measured cardiometabolic risk markers were measured by a member of the study team. Height, weight, and body composition via bioelectrical impedance (i.e., percent body fat and lean mass) were measured using a stadiometer (Seca 213, Seca, Germany) and Tanita body analyser scales (MC 780 MA, Tanita, Japan), respectively. Body mass index (BMI) was calculated and categorised into underweight (<18.5 kg/$m^2$), healthy weight (18.5-24.9 kg/$m^2$), overweight (25-29.9 kg/$m^2$) and obese (>30 kg/$m^2$) (National Health Service, 2019b).

Waist circumference (WC) was measured using the average of two measurements at the midpoint between the lowest rib and the top of the iliac crest (World Health Organization, 2011b). Blood pressure (systolic and diastolic blood pressure herein referred to as SBP & DBP) was measured using an oscillometric device after 10 minutes of seated rest (Omron 705IT, Omron, UK), and the combined highest value from both left and right grip strength measurements (Takeii analogue dynamometer, Takei, Japan) were used. A point of care device (Afinion AS100 Analyser, Alere, USA) measured glycated haemoglobin (HbA1c) levels with results classified as normoglycemic (<6.0%) or at risk (6.0-6.4%) (World Health Organization, 2011a). The submaximal modified Canadian Aerobic Fitness Test (mCAFT) estimated maximal oxygen consumption ($VO_{2max}$) (Weller et al., 1993).

## 2.4 Statistical analyses

Multi-level modelling was conducted to ascertain if there were statistically significant differences between steps, light and MVPA between devices. Then, to test the basic assumption that the selected cardiometabolic risk markers were statistically significantly associated with standard outputs on both devices, respective separate log-linear regression models were fitted to each of the markers for average daily steps, minutes of light activity and minutes of MVPA for each device. Models were adjusted for gender, age and ethnicity (white British vs. other) and assumptions were checked visually via diagnostic plots. Additionally, to establish whether the relationships between behavioural variables and each device were statistically significantly different between devices, the partial correlation of inputs and outputs were evaluated after adjustment for gender, age and ethnicity and bootstrap tested (Thulin, 2021).

A 1-out cross-validation was also performed for each model to understand how well each behavioural variable predicts each cardiometabolic marker (or predictive fit), and the resulting mean error (ME) and RMSE for the differences between observed and predicted values were calculated. The RMSE indicates how accurate the model is with smaller values corresponding to better predictive model. In

addition to single-input models and to understand the combined effect of all device variables together, multi-input models were fitted and root mean squared errors (RMSEs) compared.

Finally, to examine which behavioural variable across both devices best explain the association with the specific cardiometabolic risk markers (or explanatory fit), the proportion of variation explained by each of the models (R-squared) was reported. Akaike's information criterion corrected for small samples (AICc) was also calculated (Akaike, 1992; Hurvich & Tsai, 1989) with a smaller value of AIC corresponding to a better model (the best model for each output is calculated by $\Delta AICc = AICc - AICc_{min}$ and thus has a $\Delta$AIC value of 0), and a difference of at least 3 is considered statistically significant (Burnham et al., 2011). All analyses were conducted within R version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria).

# 2. Results

## 3.1 Descriptive statistics

Briefly, most participants had normoglycemic glucose profiles (93%), were overweight or living with obesity (88.9%) and met the UK national guidelines for physical activity (based on MVPA per day, Table 1). The number of valid days was high for both devices (6.6 days and 6.9 days for the ActiGraph and Fitbit, respectively) Participants conducted on average 6905 and 8593 steps per day as measured by the ActiGraph and Fitbit, respectively.

**Table 1.** Characteristics of the study sample

| | n | Mean (%) | SD |
|---|---|---|---|
| Age (years) | 45 | 56.0 | 8.7 |
| Sex | | | |
| Female | 27 | 60 | - |
| Male | 18 | 40 | - |
| Body mass index (kg/m$^2$) | | 31.6 | 6.9 |
| Underweight (<18.5) | 0 | 0 | - |
| Healthy weight (18.5-24.9) | 5 | 11 | - |
| Overweight (25-29.9) | 17 | 38 | - |
| Obese (>30) | 23 | 51 | - |
| Weight (kg) | 45 | 89.6 | 19.7 |
| Body fat (%) | 45 | 36.4 | 10.3 |
| HbA1c (%) | 45 | 5.6 | 0.3 |
| Normoglycemic (<6.0%) | 42 | - | - |
| At risk (6.0%–6.4%) | 3 | - | - |
| Systolic blood pressure (mm Hg) | 45 | 132.0 | 15.8 |
| Waist circumference (cm) | 45 | 101.5 | 14.8 |
| Estimated maximal oxygen uptake (ml/kg/min) | 32 | 36.7 | 6.7 |
| Grip strength (kg) | 45 | 69.1 | 22.2 |
| ActiGraph (waist-worn) | | | |
| Number of valid days | 45 | 6.6 | 0.7 |
| Steps per valid day | 45 | 6905 | 3776 |
| Light activity (mins) per valid day | 45 | 288.2 | 83.4 |
| MVPA (mins) per valid day | 45 | 33.1 | 28.4 |
| Fitbit (wrist-worn) | | | |
| Number of valid days | 45 | 6.9 | 0.3 |
| Steps per valid day | 45 | 8593 | 4543 |
| Lightly active (mins) per valid day | 45 | 231.8 | 72.5 |
| MVPA (mins) per valid day | 45 | 40.2 | 42.1 |

Notes: abbreviations (MVPA: moderate to vigorous physical activity, SD: standard deviation, min: minimum, max: maximum).

## 2.2 Movement behaviour and cardiometabolic comparisons

Results of the multi-level modelling to assess the differences of behavioural variables between devices are displayed within Table 2. The Fitbit recorded statistically significantly greater steps and fewer light minutes per day ($p < 0.001$), but there were no statistical differences when evaluating minutes of MVPA. Whilst steps from both devices had a correlation of r = 0.93, agreement varied by a large amount within individuals (65%).

**Table 2.** Comparisons between movement variables calculated via multi-level modelling.

| | Fixed effects | | | | Random effects |
|---|---|---|---|---|---|
| | Mean difference (AG-FB) | SE | *p* | r | Within individual variance (% of total) |
| Steps | -1727.1 | 264.2 | < 0.001 | 0.93 | 65% |
| Light (mins) | 50.48 | 6.8 | < 0.001 | 0.78 | 34% |
| MVPA (mins) | -6.83 | 4.17 | 0.110 | 0.67 | 31% |

Notes: The mean difference is interpretated a the ActiGraph minus Fitbit and the within individual variance has been calculated as a percent of the total variance. Abbreviations: SE (standard error), p (probability), r (Pearson's correlation coefficient), MVPA (moderate to vigorous physical activity).

Linear regression analyses confirmed that cardiometabolic risk markers were significantly associated with both devices (Table 3). Overall, step count was the most consistent predictor of cardiometabolic risk factors, with negative associations with BMI, weight, body fat and waist circumference for both the Fitbit and the ActiGraph. Despite coefficients from the ActiGraph being generally stronger, the magnitude of the logged coefficients were mostly similar between devices. Putting the coefficients into context, for every 1000 increase in step count BMI could reduce by 1.7-2.0% (keeping everything else constant). Associations for light activity and MVPA were more mixed as some cardiometabolic risk factors were related to different intensities for the Fitbit and ActiGraph monitors. Finally, analysis of partial bivariate correlations also confirmed that behavioural associations did not differ between the two devices (Supplementary Table 1).

**Table 3.** Estimated effects and respective standard errors (in brackets) of physical activity inputs on cardiometabolic risk markers, adjusting for sex, age and ethnicity.

| | INPUTS | | | | | |
|---|---|---|---|---|---|---|
| | Steps (1000s) | | Light (mins) | | MVPA (mins) | |
| OUTPUTS | FB | AG | FB | AG | FB | AG |
| BMI | 0.017 (0.006) | 0.020 (0.007) | -0.001 (0.0004) | -0.001 (0.0003) | -0.001 (0.001) | -0.002 (0.001) |
| Weight | -0.014 (0.006) | -0.017 (0.006) | -0.001 (0.0003) | -0.001 (0.0003) | -0.001 (0.001) | -0.002 (0.001) |
| Body fat | 0.021 (0.005) | 0.022 (0.006) | 0.001 (0.0003) | -0.001 (0.0003) | 0.002 (0.001) | -0.002 (0.001) |

| | | | | | | |
|---|---|---|---|---|---|---|
| SBP | 0.007 (0.004) | 0.009 (0.005) | 0.0001 (0.0003) | 0.0001 (0.0002) | 0.001 (0.001) | 0.001 (0.001) |
| DBP | 0.008 (0.004) | 0.010 (0.004) | 0.0002 (0.0002) | 0.0001 (0.0002) | 0.001 (0.001) | 0.002 (0.001) |
| HbA1c | -0.002 (0.002) | -0.002 (0.002) | -0.0001 (0.0001) | 0.000 (0.0001) | -0.0002 (0.0003) | -0.0001 (0.0003) |
| Grip strength | 0.009 (0.007) | 0.009 (0.008) | 0.001 (0.0004) | 0.001 (0.0003) | -0.001 (0.001) | 0.001 (0.001) |
| $VO_{2max}$ | 0.012 (0.005) | 0.014 (0.006) | -0.0003 (0.0004) | -0.0004 (0.0003) | 0.002 (0.001) | 0.002 (0.001) |
| WC | 0.013 (0.004) | 0.015 (0.004) | -0.001 (0.0002) | -0.001 (0.0002) | -0.001 (0.001) | -0.002 (0.001) |

Notes: The darker cell background means the effects were statistically significantly different from zero at 1% significance level. The lighter cells background corresponds to statistical significance of 5%, and the white cells are for non-statistically significant effects. Because the response variables were logged, the interpretation is on the log-scale. For example, other things being equal, an extra 1000 steps a day as measured by Fitbit is associated with an average 1.7% lower BMI. Abbreviations: MVPA (moderate to vigorous physical activity), FB (Fitbit), AG (ActiGraph), BMI (body mass index), SBP (systolic blood pressure), DBP (diastolic blood pressure), $VO_{2max}$ (estimated maximal oxygen uptake), WC (waist circumference). Results rounded to 3 decimal places where possible.

Predictive fit, i.e., the extent to which cardiometabolic risk markers can be predicted by physical activity as measured by devices, is displayed within Table 4. For single input models, the best predictive root mean squared error (RMSE) ranged from 0.0639 for HbA1c (i.e., a 95% predictive interval of ± 6.6%) to 0.1853 (i.e., a 95% predictive interval of ± 20%) for grip strength. No device was found to provide a consistently better prediction across the cardiometabolic risk markers. Combining the inputs into one model generally resulted in a smaller RMSE with the exception of DBP, SBP and HBA1c.

**Table 4.** Comparing how well each behavioural variable predicts each cardiometabolic marker (or predictive fit) in terms of mean error (ME) and root mean squared error (RMSE, in brackets below) for the models with individual Fitbit and ActiGraph inputs and all inputs simultaneously.

| | INPUTS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Steps (1000s) | | Light (mins) | | MVPA (mins) | | All | All |
| OUTPUTS | FB | AG | FB | AG | FB | AG | FB | AG |
| BMI | -0.001 (0.175) | -0.002 (0.178) | 0.001 (0.173) | 0.002 (0.182) | 0.001 (0.186) | -0.001 (0.183) | 0.001 (0.169) | 0.002 (0.120) |
| Weight | -0.0003 (0.162) | 0.001 (0.161) | 0.001 (0.165) | 0.002 (0.170) | 0.001 (0.171) | -0.0001 (0.165) | -0.002 (0.177) | 0.003 (0.074) |
| Body fat | 0.002 (0.152) | 0.001 (0.157) | 0.002 (0.166) | 0.003 (0.175) | 0.003 (0.169) | 0.004 (0.175) | 0.004 (0.165) | 0.002 (0.076) |
| SBP | 0.001 (0.125) | 0.001 (0.123) | 0.002 (0.130) | 0.001 (0.130) | 0.001 (0.125) | 0.001 (0.124) | 0.002 (0.166) | 0.006 (0.206) |
| DBP | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.007 | -0.001 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (0.117) | (0.116) | (0.123) | (0.124) | (0.121) | (0.115) | (0.159) | (0.192) |
| HbA1c | -0.001 | -0.001 | 0.0001 | 0.001 | -0.0004 | -0.001 | 0.004 | 0.014 |
| | (0.066) | (0.067) | (0.064) | (0.066) | (0.067) | (0.067) | (0.146) | (0.168) |
| Grip strength | 0.003 | 0.004 | 0.0001 | 0.0003 | 0.003 | 0.002 | 0.001 | 0.010 |
| | (0.197) | (0.199) | (0.195) | (0.185) | (0.197) | (0.199) | (0.132) | (0.152) |
| $VO_{2max}$ | -0.002 | 0.001 | -0.003 | -0.002 | -0.002 | 0.0001 | -0.0004 | 0.005 |
| | (0.125) | (0.122) | (0.146) | (0.143) | (0.126) | (0.123) | (0.130) | (0.115) |
| WC | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.005 | 0.005 |
| | (0.112) | (0.113) | (0.118) | (0.121) | (0.122) | (0.119) | (0.122) | (0.115) |

Notes: Mean errors (MEs) closer to zero indicate less bias and smaller root mean squared errors (RMSEs) indicate higher precision of prediction. The shaded cells indicate the models with the lowest RMSE values for each output. Note, that since the response was logged, the results are to be interpreted on a logarithmic scale. Thus, for example, predictions of systolic blood pressure based on FB steps alone will on average be off by 0.09%, and 95% of them will fall within 26.6% of the true value. Abbreviations: MVPA (moderate to vigorous physical activity), FB (Fitbit), AG (ActiGraph), BMI (body mass index), SBP (systolic blood pressure), DBP (diastolic blood pressure), $VO_{2max}$ (estimated maximal oxygen uptake), WC (waist circumference). The darker cell background highlights the best RMSE pairing for each cardiometabolic marker, for both individual and grouped input models. Results rounded to 3 decimal places where possible.

Analyses to understand the extent to which each behavioural variable from both devices explains the specific cardiometabolic risk markers (explanatory fit) were also conducted (Supplementary Table 2). This showed that whilst the number of steps as recorded by ActiGraph was the best explanatory variable for the majority of cardiometabolic variables including BMI, weight, SBP, $V0_{2max}$ and WC, Fitbit models were not statistically significantly worse for those cardiometabolic risk markers (AICc < 3).

# 3. Discussion

This study confirms that when using step count data derived from wearable devices to predict cardiometabolic risk markers, relationships are similar irrespective of whether a research grade or a commercial grade device was been used. Whilst data from the waist-worn ActiGraph produced stronger estimates and better explanatory fit, models predicting cardiometabolic risk markers from step data were not statistically different between devices (BMI, weight, body fat %, SBP, DBP, $VO_{2max}$ and WC AICc differences < 3). These results indicate that the predictive ability of a commercial wrist-worn wearable device was similar, and provides a rationale for their use in scenarios that research grade devices are typically used.

Our finding that body composition variables are associated with movement behaviours derived from wearable devices is consistent with the existing literature (Lim et al., 2018; Rykov et al., 2020). However, in contrast to the current study, Rykov et al. (2020) demonstrated that only blood based markers (in their case HDL and Triglycerides) were significantly associated with variables derived from steps and not body composition. There is a lack of supporting literature within this area, likely due to the unpractical deployment of two devices into health screening interventions. Despite this, the significant associations shown in this study point to the potential beneficial application of commercial wearables within the health and wellbeing sector. Yet, the proprietary nature of how commercial wearable devices calculate physical activity metrics and the variation in accelerometric signal transformation into behavioural outcome variables are key concerns for using wearables within health research (Troiano et al., 2020). Within this study, steps were shown to be one of the better input variables for explanatory fit and although minutes of behaviour are often used to compare national guideline compliance, steps are often the most easily understood and widely used metric and could suffer from less proprietary processing.

Population level surveillance of physical activity using wearable devices would provide an understanding to the level of compliance with national guidelines but there are significant barriers for wide scale adoption. Research grade devices are often costly and are not designed for longitudinal monitoring unlike consumer developed monitors. The aim of this study was to assess if the outputs from commercial devices are comparable to a research grade device, and we have shown that for some cardiometabolic variables coefficients were similar. Based on our models, for every 1000 increase in steps as measured by the Fitbit could result in an average 1.7% decrease in body mass index or 1.4% decrease in body weight. With consumer wearable interventions on average increasing physical activity by 2,123 steps per day (Franssen et al., 2020), if an individual with the average BMI of the sample (31.82 $kg/m^2$) did 2500 more steps per day, BMI could drop to by 1.35 $kg/m^2$ to 30.46 $kg/m^2$. Therefore, linking metrics provided by commercial wearable devices into existing healthcare pathways such as patient medical records could provide the necessary platform for prevention-focused activities, and offer an opportunity for important conversations about health and wellbeing within routine interactions with health professionals. The integration of digital health technologies

within healthcare systems is on the agenda of the UK government to support future disease prevention pathways (National Health Service, 2019a; Taskforce on Innovation, 2021) but significant concerns include representativeness, data ownership and longevity of devices (Troiano et al., 2020).

As only one of the few studies that have investigated the link between physical activity behaviours measured by wearable devices and cardiometabolic risk markers, there is a need for more research to investigate the opportunity that wearable devices hold for healthcare. This is highlighted by variability in the literature that suggests wearable interventions can provide between -0.4 to -4.4 kg and 0.08 to -3.43 kg/$m^2$ changes in body weight and BMI, respectively (McDonough et al., 2021). Greater population level data are therefore required to understand if the associations presented here under or overestimate the relationship with body composition variables, and if larger datasets can explain associations with other important cardiometabolic risk markers.

Although this study is novel and to our knowledge the first to compare behavioural estimates from two devices to the same cardiometabolic risk markers, the sample size is small which limits the generalizability of the data. However, our work does offer proof of concept insights into the potential use of commercial wearables as a substitute for research grade devices. The exclusion of wear time from the statistical models was a limitation as it prevented us from determining if volumetric discrepancies were device or wear related. This approach was taken due to a lack of standardised methods for deciphering wear time and sedentary time within Fitbit devices. Additionally, the sedentary time contamination prevented the use of modelling techniques that accounts for the co-dependence of behaviours (compositional data analyses) and should be explored in future datasets that has dual deployment of devices. The difference in wear sites (waist and wrist) between devices must also be acknowledged, which could explain some of the variation in models in intensity variables. Our study also focused on a device that at the time was popular and available but now is not on the commercial market due to the release of more recent models. However, this approach could and should be taken with a larger dataset and with the latest emerging devices.

# 4. Conclusion

Regardless of whether step count data from a research grade or commercial grade device was used, similar associations were found for BMI, weight, body fat % and WC. Overall, these results suggest that the predictive and explanatory ability of step count data from both devices with selected cardiometabolic risk markers is similar, which may offer additional opportunities for health research from commercial wearables. Further work exploring the dual deployment of both types of devices are required to confirm these findings.

# Supplementary material

Analysis of partial bivariate correlations revealed that behavioural associations did not differ between devices as Fitbit inputs were not statistically significantly different from those for AG inputs (Supplementary Table 1).

**Supplementary Table 1.** Partial correlations (after adjusting for demographic variables) between individual inputs and outputs to test the hypothesis that the differences between partial correlations for Fitbit and ActiGraph inputs are statistically significant.

| | INPUTS | | | | | |
|---|---|---|---|---|---|---|
| | Steps (1000s) | | Light (mins) | | MVPA (mins) | |
| OUTPUTS | FB | AG | FB | AG | FB | AG |
| BMI | -0.40 (0.49) | -0.42 | -0.38 (0.38) | -0.30 | -0.19 (0.24) | -0.32 |
| Weight | -0.36 (0.37) | -0.39 | -0.31 (0.65) | -0.25 | -0.18 (0.18) | -0.32 |
| Body fat | -0.56 (0.88) | -0.54 | -0.42 (0.31) | -0.32 | -0.39 (0.83) | -0.36 |
| SBP | 0.26 (0.80) | 0.28 | 0.07 (0.97) | 0.05 | 0.26 (0.95) | 0.27 |
| DBP | 0.33 (0.75) | 0.34 | 0.13 (0.69) | 0.05 | 0.22 (0.09) | 0.38 |
| HbA1c | -0.14 (0.77) | -0.10 | -0.16 (0.70) | -0.06 | -0.12 (0.86) | -0.07 |
| Grip strength | 0.21 (0.75) | 0.19 | 0.22 (0.17) | 0.36 | 0.20 (0.45) | 0.09 |
| $VO_{2max}$ | 0.36 (0.32) | 0.41 | -0.14 (0.56) | -0.23 | 0.35 (0.70) | 0.39 |
| WC | -0.47 (0.64) | -0.48 | -0.36 (0.72) | -0.33 | -0.30 (0.34) | -0.38 |

Notes: The bootstrapped p-values (in brackets) are for testing the hypothesis that the differences between partial correlations for Fitbit input and for ActiGraph inputs are statistically significant. Abbreviations: MVPA (moderate to vigorous physical activity), FB (Fitbit), AG (ActiGraph), BMI (body mass index), SBP (systolic blood pressure), DBP (diastolic blood pressure), $VO_{2max}$ (estimated maximal oxygen uptake), WC (waist circumference). Results rounded to 2 decimal places.

The number of steps as recorded by ActiGraph was the best explanatory variable for BMI ($R^2$ = 0.42), weight ($R^2$ = 0.59), WC ($R^2$=0.58), SBP ($R^2$=0.25) and $V0_{2max}$ ($R^2$=0.66). For body fat, DBP, and HBA1c, the best explanatory variables were the number of steps as recorded by FitBit ($R^2$=0.81), the MVPA as recorded by ActiGraph ($R^2$=0.46), and the minutes of light activity as recorded by FitBit ($R^2$=0.10) respectively. However, none of those were statistically significantly worse than the number of steps as recorded by ActiGraph (ΔAICc=1.9, ΔAICc=1.5, and ΔAICc=0.8 respectively). The only cardiometabolic marker for which the model with the number of steps as recorded by ActiGraph as an explanatory variable was statistically significantly worse than the best model (the minutes of light activity as recorded by ActiGraph, $R^2$=0.75) was Grip strength (ΔAICc=4.8).

**Supplementary Table 2.** The extent to which each behavioural variable, as measured by both devices, explains specific cardiometabolic risk markers (explanatory fit) expressed via corrected Akaike Information Criterion (AICc) and R-squared (in brackets).

| | INPUTS | | | | | |
|---|---|---|---|---|---|---|
| | Steps (1000s) | | Light (mins) | | MVPA (mins) | |
| OUTPUTS | FB | AG | FB | AG | FB | AG |
| BMI | 0.9 (0.41) | **0.0** (0.42) | 1.3 (0.40) | 4.6 (0.35) | 7.1 (0.32) | 4.0 (0.36) |
| Weight | 1.2 (0.58) | **0.0** (0.59) | 3.0 (0.56) | 4.6 (0.54) | 6.1 (0.53) | 2.8 (0.56) |
| Body fat | **0.0** (0.81) | 1.9 (0.81) | 8.2 (0.78) | 12.4 (0.75) | 9.4 (0.77) | 11.1 (0.76) |
| SBP | 0.6 (0.24) | **0.0** (0.25) | 3.7 (0.19) | 3.7 (0.19) | 0.6 (0.24) | 0.4 (0.25) |
| DBP | 2.0 (0.43) | 1.5 (0.44) | 6.4 (0.38) | 7.1 (0.37) | 4.8 (0.40) | **0.0** (0.46) |
| HbA1c | 0.3 (0.09) | 0.8 (0.08) | **0.0** (0.10) | 1.1 (0.08) | 0.5 (0.09) | 1.0 (0.08) |
| Grip strength | 4.4 (0.73) | 4.8 (0.72) | 4.1 (0.73) | **0.0** (0.75) | 4.6 (0.73) | 6.1 (0.72) |
| $VO_{2max}$ | 1.2 (0.64) | **0.0** (0.66) | 5.9 (0.59) | 4.9 (0.60) | 1.4 (0.64) | 0.2 (0.66) |
| WC | 0.6 (0.57) | **0.0** (0.58) | 5.5 (0.52) | 7.0 (0.50) | 7.8 (0.50) | 5.3 (0.52) |

Notes: For the ease of comparison, for each output, the difference between the AICc for the model with each specified input and the best model for that output is shown. The best model for each output thus has a ΔAICc value of 0 (highlighted by the darker cell background). AICc allows for statistical comparison of non-nested models with the same input with the difference of at least 3 is required for statistical significance. Abbreviations: MVPA (moderate to vigorous physical activity), FB (Fitbit), AG (ActiGraph), BMI (body mass index), SBP (systolic blood pressure), DBP (diastolic blood pressure), $VO_{2max}$ (estimated maximal oxygen uptake), WC (waist circumference). R-squared rounded to 2 decimal places.

# References


Akaike, H. (1992). *Information Theory and an Extension of the Maximum Likelihood Principle* (pp. 610–624). https://doi.org/10.1007/978-1-4612-0919-5_38

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35. https://doi.org/10.1007/s00265-010-1029-6

Chastin, S., McGregor, D., Palarea-Albaladejo, J., Diaz, K. M., Hagströmer, M., Hallal, P. C., van Hees, V. T., Hooker, S., Howard, V. J., Lee, I.-M., von Rosen, P., Sabia, S., Shiroma, E. J., Yerramalla, M. S., & Dall, P. (2021). Joint association between accelerometry-measured daily combination of time spent in physical activity, sedentary behaviour and sleep and all-cause mortality: a pooled analysis of six prospective cohorts using compositional analysis. *British Journal of Sports Medicine*, bjsports-2020-102345. https://doi.org/10.1136/bjsports-2020-102345

Chu, A. H. Y., Ng, S. H. X., Paknezhad, M., Gauterin, A., Koh, D., Brown, M. S., & Müller-Riemenschneider, F. (2017). Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLOS ONE*, *12*(2), e0172535. https://doi.org/10.1371/journal.pone.0172535

Esliger, D. W., Kingsnorth, A. P., & Sherar, L. B. (2017). Measuring Physical Activity. In J. Piggin, L. Mansfield, & M. Weed (Eds.), *Routledge handbook of physical activity policy and practice* (1st Editio). Routledge.

Franssen, W. M. A., Franssen, G. H. L. M., Spaas, J., Solmi, F., & Eijnde, B. O. (2020). Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. *International Journal of Behavioral Nutrition and Physical Activity*, *17*(1), 57. https://doi.org/10.1186/s12966-020-00955-2

Gray, L. J., Taub, N. A., Khunti, K., Gardiner, E., Hiles, S., Webb, D. R., Srinivasan, B. T., & Davies, M. J. (2010). The Leicester Risk Assessment score for detecting undiagnosed Type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. *Diabetic Medicine*, *27*(8), 887–895. https://doi.org/10.1111/j.1464-5491.2010.03037.x

Hajna, S., Ross, N. A., & Dasgupta, K. (2018). Steps, moderate-to-vigorous physical activity, and cardiometabolic profiles. *Preventive Medicine*, *107*, 69–74. https://doi.org/10.1016/j.ypmed.2017.11.007

Huang, B.-H., Inan-Eroglu, E., Hamer, M., & Stamatakis, E. (2020). Joint associations of device-measured physical activity and sleep duration with cardiometabolic health in the 1970 British Cohort Study. *Journal of Science and Medicine in Sport*, *23*(12), 1191–1196. https://doi.org/10.1016/j.jsams.2020.07.012

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Janssen, I., Clarke, A. E., Carson, V., Chaput, J.-P., Giangregorio, L. M., Kho, M. E., Poitras, V. J., Ross, R., Saunders, T. J., Ross-White, A., & Chastin, S. F. M. (2020). A systematic review of

compositional data analysis studies examining associations between sleep, sedentary behaviour, and physical activity with health outcomes in adults. *Applied Physiology, Nutrition, and Metabolism*, *45*(10 (Suppl. 2)), S248–S257. https://doi.org/10.1139/apnm-2020-0160

Kingsnorth, A. P., Patience, M., Moltchanova, E., Esliger, D. W., Paine, N. J., & Hobbs, M. (2021). Changes in Device-Measured Physical Activity Patterns in U.K. Adults Related to the First COVID-19 Lockdown. *Journal for the Measurement of Physical Behaviour*, 1–10. https://doi.org/10.1123/jmpb.2021-0005

Lim, W. K., Davila, S., Teo, J. X., Yang, C., Pua, C. J., Blöcker, C., Lim, J. Q., Ching, J., Yap, J. J. L., Tan, S. Y., Sahlén, A., Chin, C. W.-L., Teh, B. T., Rozen, S. G., Cook, S. A., Yeo, K. K., & Tan, P. (2018). Beyond fitness tracking: The use of consumer-grade wearable data from normal volunteers in cardiovascular and lipidomics research. *PLOS Biology*, *16*(2), e2004285. https://doi.org/10.1371/journal.pbio.2004285

McDonough, D. J., Su, X., & Gao, Z. (2021). Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis. *British Journal of Sports Medicine*, bjsports-2020-103594. https://doi.org/10.1136/bjsports-2020-103594

Mikkelsen, M.-L. K., Berg-Beckhoff, G., Frederiksen, P., Horgan, G., O'Driscoll, R., Palmeira, A. L., Scott, S. E., Stubbs, J., Heitmann, B. L., & Larsen, S. C. (2020). Estimating physical activity and sedentary behaviour in a free-living environment: A comparative study between Fitbit Charge 2 and Actigraph GT3X. *PLOS ONE*, *15*(6), e0234426. https://doi.org/10.1371/journal.pone.0234426

National Health Service. (2019a). *NHS Long Term Plan » Online version of the NHS Long Term Plan*. https://www.longtermplan.nhs.uk/online-version/

National Health Service. (2019b). *What is the body mass index (BMI)?* Department of Health. https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/

OFCOM. (2021). *OFCOM Technology Tracker 2021*. https://www.ofcom.org.uk/__data/assets/pdf_file/0015/219102/technology-tracker-2021-data-tables.pdf

Pontin, F., Lomax, N., Clarke, G., & Morris, M. A. (2021). Socio-demographic determinants of physical activity and app usage from smartphone data. *Social Science & Medicine*, *284*, 114235. https://doi.org/10.1016/j.socscimed.2021.114235

Rykov, Y., Thach, T.-Q., Dunleavy, G., Roberts, A. C., Christopoulos, G., Soh, C. K., & Car, J. (2020). Activity Tracker–Based Metrics as Digital Markers of Cardiometabolic Health: Cross-Sectional Study. *JMIR MHealth and UHealth*, *8*(1), e16409. https://doi.org/10.2196/16409

Strain, T., Wijndaele, K., & Brage, S. (2019). Physical Activity Surveillance Through Smartphone Apps and Wearable Trackers: Examining the UK Potential for Nationally Representative Sampling. *JMIR MHealth and UHealth*, *7*(1), e11898. https://doi.org/10.2196/11898

Taskforce on Innovation, G. and R. R. (2021). *Independent report*. https://www.gov.uk/government/publications/taskforce-on-innovation-growth-and-regulatory-reform-independent-report

Teo, J. X., Davila, S., Yang, C., Hii, A. A., Pua, C. J., Yap, J., Tan, S. Y., Sahlén, A., Chin, C. W.-L., Teh, B. T., Rozen, S. G., Cook, S. A., Yeo, K. K., Tan, P., & Lim, W. K. (2019). Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. *Communications Biology*, *2*(1), 361. https://doi.org/10.1038/s42003-019-0605-1

Thulin, M. (2021). *Modern statistics with R*. Eos Chasma Press. http://www.modernstatisticswithr.com/

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3

Troiano, R. P., Stamatakis, E., & Bull, F. C. (2020). How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions. *British Journal of Sports Medicine*, *54*(24), 1468–1473. https://doi.org/10.1136/bjsports-2020-102621

Tudor-Locke, C., Barreira, T. V., & Schuna, J. M. (2015). Comparison of Step Outputs for Waist and Wrist Accelerometer Attachment Sites. *Medicine & Science in Sports & Exercise*, *47*(4), 839–842. https://doi.org/10.1249/MSS.0000000000000476

Van Blarigan, E. L., Kenfield, S. A., Tantum, L., Cadmus-Bertram, L. A., Carroll, P. R., & Chan, J. M. (2017). The Fitbit One Physical Activity Tracker in Men With Prostate Cancer: Validation Study. *JMIR Cancer*, *3*(1), e5. https://doi.org/10.2196/cancer.6935

Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, *174*(6), 801–809. https://doi.org/10.1503/cmaj.051351

Weller, I. M., Thomas, S. G., Corey, P. N., & Cox, M. H. (1993). Prediction of maximal oxygen uptake from a modified Canadian aerobic fitness test. *Canadian Journal of Applied Physiology*, *18*(2), 175–188. https://doi.org/10.1139/h93-014

Whelan, M. E., Denton, F., Bourne, C. L. A., Kingsnorth, A. P., Sherar, L. B., Orme, M. W., & Esliger, D. W. (2021). A digital lifestyle behaviour change intervention for the prevention of type 2 diabetes: a qualitative study exploring intuitive engagement with real-time glucose and physical activity feedback. *BMC Public Health*, *21*(1), 130. https://doi.org/10.1186/s12889-020-09740-z

Whelan, M. E., Kingsnorth, A. P., Orme, M. W., Sherar, L. B., & Esliger, D. W. (2017). Sensing interstitial glucose to nudge active lifestyles (SIGNAL): feasibility of combining novel self-monitoring technologies for persuasive behaviour change. *BMJ Open*, *7*(10), e018282. https://doi.org/10.1136/bmjopen-2017-018282

World Health Organization. (2011a). *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus*.

World Health Organization. (2011b). Waist Circumference and Waist–Hip Ratio. *WHO Expert*, *64*(1), 2–5. http://www.nature.com/doifinder/10.1038/ejcn.2009.139