Cardiorespiratory fitness estimation using wearable sensors: laboratory and free-living analysis of context-specific submaximal heart rates

J Appl Physiol 120: 1082–1096, 2016.
First published March 3, 2016; doi:10.1152/japplphysiol.00519.2015.

Cardiorespiratory fitness estimation using wearable sensors: Laboratory and free-living analysis of context-specific submaximal heart rates

Marco Altini,[1] Pierluigi Casale,[2] Julien Penders,[2] Gabrielle ten Velde,[3] Guy Plasqui,[3] and Oliver Amft[4,5]

[1]*Eindhoven University of Technology, The Netherlands and Bloom Technologies, Diepenbeek, Belgium;* [2]*Holst Centre/imec, Eindhoven, The Netherlands;* [3]*Human Biology Department, Maastricht University, Maastricht, The Netherlands; and* [4]*University of Passau, Passau, Germany; and* [5]*Eindhoven University of Technology, Eindhoven, The Netherlands*

Submitted 18 June 2015; accepted in final form 28 February 2016

Altini M, Casale P, Penders J, ten Velde G, Plasqui G, Amft O. Cardiorespiratory fitness estimation using wearable sensors: Laboratory and free-living analysis of context-specific submaximal heart rates. *J Appl Physiol* 120: 1082–1096, 2016. First published March 3, 2016; doi:10.1152/japplphysiol.00519.2015.—In this work, we propose to use pattern recognition methods to determine submaximal heart rate (HR) during specific contexts, such as walking at a certain speed, using wearable sensors in free living, and using context-specific HR to estimate cardiorespiratory fitness (CRF). CRF of 51 participants was assessed by a maximal exertion test ($\dot{V}O_{2\,max}$). Participants wore a combined accelerometer and HR monitor during a laboratory-based simulation of activities of daily living and for 2 wk in free living. Anthropometrics, HR while lying down, and walking at predefined speeds in laboratory settings were used to estimate CRF. Explained variance (R^2) was 0.64 for anthropometrics, and increased up to 0.74 for context-specific HR (0.73-0.78 when including fat-free mass). Next, we developed activity recognition and walking speed estimation algorithms to determine the same contexts (i.e., lying down and walking) in free living. Context-specific HR in free living was highly correlated with laboratory measurements (Pearson's $r = 0.71$–0.75). R^2 for CRF estimation was 0.65 when anthropometrics were used as predictors, and increased up to 0.77 when including free-living context-specific HR (i.e., HR while walking at 5.5 km/h). R^2 varied between 0.73 and 0.80 when including fat-free mass among the predictors. Root mean-square error was reduced from 354.7 to 281.0 ml/min by the inclusion of context-specific HR parameters (21% error reduction). We conclude that pattern recognition techniques can be used to contextualize HR in free living and estimated CRF with accuracy comparable to what can be obtained with laboratory measurements of HR response to walking.

cardiorespiratory fitness; wearable sensors; heart rate; physical activity; context recognition

NEW & NOTEWORTHY

Many methods have been developed to estimate $\dot{V}O_{2\,max}$ using data collected under supervised laboratory conditions or following strict protocols. However, to the best of our knowledge, this is the first work that proposes pattern recognition methods to contextualize heart rate in free living and use context-specific heart rate to predict $\dot{V}O_{2\,max}$. The proposed method does not require laboratory tests or specific protocols, showing error reductions up to 21% compared with $\dot{V}O_{2\,max}$ estimates derived using anthropometrics only.

CARDIORESPIRATORY FITNESS (CRF) is a diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals, and can be adopted as a proxy of cardiovascular and cardiorespiratory health (16, 26). Thus, CRF is a marker of training status that can be considered one of the most important determinants of health and wellbeing. While recent developments in wearable sensor technologies improved the accuracy of physical activity monitoring devices in daily life, almost all solutions focus on behavioral aspects such as steps, activity type, and energy expenditure (EE) (4, 6). Steps or EE are relevant markers of an individual's health; however, they mainly reflect the individual's behavior instead of the individual's health status. CRF estimation using wearable sensors could provide more insight on an individual's health status, noninvasively, and therefore help clinicians and individuals coaching or leading a healthier lifestyle.

Currently, the gold standard for CRF measurement is performed by direct measurement of oxygen consumption during maximal exercise (i.e., $\dot{V}O_{2\,max}$) (30, 31). However, $\dot{V}O_{2\,max}$ measurements require medical supervision and can be risky for individuals where exercise until maximal exertion is contraindicated. Despite the indubitable importance of CRF in health, measurements of $\dot{V}O_{2\,max}$ are therefore rare (21), and less risky submaximal tests have been developed. Nonexercise CRF estimation models use easily accessible measures such as age, gender, and a self-reported physical activity level (12, 20). However, for individuals with similar anthropometric characteristics, CRF levels cannot be discriminated accurately. Alternatively, submaximal tests have been introduced to estimate $\dot{V}O_{2\,max}$ during specific protocols while monitoring heart rate (HR) at predefined workloads (5, 10). The strict workload imposed by the protocol is used to exploit the inverse relation between HR in a specific context (e.g., while running or biking at a specific intensity) and $\dot{V}O_{2\,max}$. However, the need for laboratory equipment and the necessity to reperform the test to detect changes in CRF limit the practical applicability of such techniques. Ideally, we would like to estimate CRF in free living during activities of daily living, thus without the need for specific laboratory tests or exercise protocols. Estimating CRF using wearable sensor data acquired during regular activities of daily living could provide continuous assessment without the need for specific tests or protocols.

Miniaturized wearable sensors combining accelerometer and HR data provide a way to investigate the relation between physical activity, HR, and $\dot{V}O_{2\,max}$ in free living. Additionally, advances in signal processing and machine learning techniques recently provided new methods to accurately recognize contexts in which HR can be analyzed, such as activity type, walking speed, and EE (1, 6, 28), in free living.

The relation between submaximal HR during activities of daily living simulated in laboratory settings and $\dot{V}O_{2\,max}$ has been evaluated by different research groups (2, 23, 28, 30).

Address for reprint requests and other correspondence: M. Altini, Eindhoven Univ. of Technology, Den Dolech 2, 5612AZ (e-mail: altini.marco@gmail.com).

Table 1. *Participants' characteristics*

Parameter	
n	51 (24 male, 27 female)
Age, yr	25.1 ± 6.0
Body weight, kg	68.4 ± 10.8
BMI, kg/m^2	22.7 ± 2.5
Fat-free mass, kg	52.6 ± 9.2
$\dot{V}O_{2max}$, ml/min	3,037.5 ± 671.6

Data are means ± SD; *n*, no. of subjects. BMI, body mass index.

Tonis et al. (29) explored different parameters to estimate CRF from HR and accelerometer data during activities of daily living simulated in laboratory settings. However, $\dot{V}O_{2\,max}$ reference and free-living data were not collected. Others (2, 7, 23) measured HR parameters representative of CRF in the context of improving estimates of energy expenditure, showing how interindividual differences in HR could be accounted for by surrogates of fitness such as measured or estimated submaximal HR. In free-living conditions, the relation between physical activity as expressed by a step counter and CRF was investigated by Cao et al. (9). While steps could provide useful insights, the relation between HR and $\dot{V}O_2$ at a certain exercise intensity cannot be exploited using only motion-based sensors. Plasqui and Westerterp (22) showed that a combination of average HR and physical activity over a period of 7 days correlates significantly with $\dot{V}O_{2\,max}$. However, the relation between average HR and activity counts depends on the amount of activity performed, and therefore could also be affected by behavioral correlates of CRF. Many studies showed strong links between submaximal HR during simulated activities of daily living and CRF, thus motivating our research.

In this study, we aimed at investigating the relation between submaximal HR in specific contexts as recorded by wearable sensors in free living, and CRF, and to predict $\dot{V}O_{2\,max}$ using free-living data. To this aim, we hypothesized that isolating the same contexts in laboratory settings and in free living using pattern recognition methods could yield similar relations between context-specific HR and $\dot{V}O_{2\,max}$.

METHODS

Participants

Participants were 51 (24 male, 27 female) healthy adults. Anthropometric characteristics and CRF level are reported in Table 1. Written informed consent was obtained by each participant. The study was approved by the medical ethics committee of Maastricht University.

ECG and Accelerometer Device

The sensor platform used was an ECG Necklace. The ECG Necklace (1) is a low-power wireless ECG platform. The system relies on an ultra-low-power ASIC for ECG readout, and it is integrated in a necklace, providing ease-of-use and comfort while allowing flexibility in lead positioning and system functionality. It achieves up to 6 days autonomy on a 175-mAh lithium-ion battery. For the current study, the ECG Necklace was configured to acquire one-lead ECG data at 256 Hz and accelerometer data from a triaxial accelerometer (ADXL330) at 64 Hz. The ADXL330 accelerometer provides a ±3-g range and high sensitivity (300 mV/g) and was digitalized to 12 bits input by the ECG Necklace. The *x*-, *y*-, and *z*-axes of the accelerometer were oriented along the vertical, mediolateral, and anteroposterior directions of the body, respectively. The ECG Necklace was not attached to the body to improve user comfort during free living. Two gel electrodes were placed on the participant's chest in the lead II configuration. Data were recorded on the on-board SD card to ensure integrity.

The ECG Necklace was previously validated as a reliable physical activity monitor able to quantify different physical activity parameters with high accuracy, such as activity type, walking speed, and EE (1, 3). A continuous wavelet transform-based beat detection algorithm was used to extract RR intervals from ECG data (24). Segments of data identified as lying or sedentary (no or limited movement) as well as flat ECG signal or inaccurate HR were treated as "monitor not worn." Inaccurate HR was identified as periods where consecutive RR intervals varied >20%, as typically performed in clinical practice for HR variability analysis.

Indirect Calorimetry

The gas analysis was performed with an open-circuit indirect calorimeter in diluted flow mode, meaning that the subject could freely breathe in an airstream. The flow past the subject mouth was set at 400 l/min. This means subject breathing ventilation of up to 200 l/min can be measured without rebreathing except for the volume of the applied face mask. Total flow was measured and converted to standard temperature and pressure, dry values with a large dry-bellows flowmeter calibrated to 0.2% of used range by the national standards bureau (5-point calibration) and using calibrated temperature, humidity, and pressure sensors. Gas samples taken from the flow are filtered, dried, pressurized, and fed into high-resolution O_2 and CO_2 analyzers made by ABB-Hartmann & Braun (OA2020 and Easyline 19-in. rack units) and Servomex (Servomex 4100 and Servopro 5400 19-in. rack units) with a resolution ≤0.001% absolute. The analyzers were mounted separately to exclude both vibration and climate variation as confounding factors. Ranges for the analyzers were set to 0–21% O_2 and 0–1% CO_2 yet only limited to 25 and 2.5%, respectively. No specific smoothing was applied, since the result was updated each 5 s while breathing was mechanically averaged in the ±30 liter internal volume and the dilution gas flow, resulting in a time constant of 4.5 s at the 400 l/min setting. The calorimeter was validated by gas infusion or burning fuel (methanol) over its full range (200-7,000 ml/min) with a 1 ± 2% average ± SD result. $\dot{V}O_{2\,max}$ was reached when a plateau in $\dot{V}O_2$ was observed and/or at a respiratory quotient of 1.1 or higher. $\dot{V}O_{2\,max}$ was calculated as the highest average $\dot{V}O_2$ over 30 s (6 consecutive values).

Study Design

The ECG Necklace was worn during laboratory protocols and free living.

Laboratory protocols. Participants reported at the laboratory on three separate days and after refraining from drinking, eating, and smoking in the 2 h before the experiment. Two laboratory protocols were performed, while the 3rd day was used for anthropometric measurements, including the participant's body weight, height, and body fat.

The first protocol included simulated activities performed while connected to an indirect calorimeter (Omnical; Maastricht University, Maastricht, The Netherlands) to determine context-specific HR during activities of daily living simulated in laboratory settings. Activities included lying down, sitting, sitting and writing, standing, cleaning a table, sweeping the floor, walking (treadmill flat at 2.5, 3, 3.5, 4, 4.5, 5, 5.5, and 6 km/h), and running (treadmill flat at 7, 8, 9, and 10 km/h). Activities were carried out for a period of at least 4 min.

The second protocol was a $\dot{V}O_{2\,max}$ test providing reference data for biking and CRF. $\dot{V}O_{2\,max}$ was determined during an incremental test on a cycle ergometer according to the protocol of Kuipers et al. (15). After a 5-min warm-up at 100 W for men and 75 W for women,

workload was increased by 50 W every 2.5 min. When the HR reached 35 beats/min below the age-predicted maximal HR (208-0.7 × age) or the respiratory quotient exceeded one, workload was increased by 25 W every 2.5 min until exhaustion. Expired air was continuously analyzed for O_2 consumption and CO_2 production using indirect calorimetry.

Free-living protocol. Participants wore the ECG Necklace for 14 consecutive days in free living while carrying out their normal activities of daily living. Participants were instructed to wear the ECG Necklace during day and night, except during showering, water activities, or charging of the ECG Necklace, since the ECG Necklace is not waterproof. Charging was performed daily for 1 h. Participants were also instructed to change electrodes daily or after physical exercise.

Data Processing

Context-specific HR in laboratory settings was determined as the mean HR during scripted activities performed by the participant and combined with anthropometrics in a regression model to predict $\dot{V}O_{2\,max}$. The regression model was analyzed to validate the assumption that submaximal context-specific HR can be used to estimate CRF level. Activity type recognition and walking speed models were built using data from laboratory settings and used in free living. For each participant, models were built using only data from other participants. Therefore, all models were nonindividualized, and no laboratory data from the participant to be validated were used for model building. The procedure used for model building and evaluation is shown in Fig. 1. For the beat detection we relied on methods developed by the research community in the past, since these models are standard components that are already available in many sensor devices today. More details on the validation procedures are reported in *Statistics*. Context-specific HR in free living was used in a multiple-regression model to estimate $\dot{V}O_{2\,max}$ without the need for laboratory protocols and analyzed with respect to results obtained using submaximal context-specific HR acquired during activities of daily living simulated in laboratory settings.

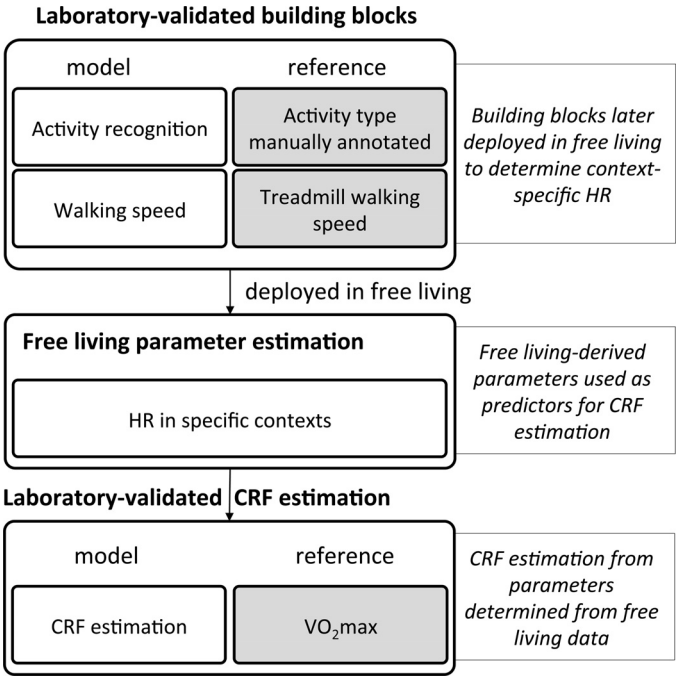

Fig. 1. Block diagram of the proposed approach and validation procedure. Activity recognition and walking speed estimation models are built and validated using supervised laboratory recordings. Next, models are deployed in free living. Activity recognition and walking speed estimation are used to determine HR in specific contexts in free living. Finally, HR in specific contexts [e.g., heart rate (HR) while lying down or walking at a certain speed] are used as predictors for $\dot{V}O_{2\,max}$ estimation, effectively estimating cardiorespiratory fitness (CRF) level from free-living data. All models are validated using leave-one-subject-out cross-validation, i.e., no data used for model validation were used for model building, as described in *Statistics*. An example of activity recognition and walking speed estimation models output is shown in Fig. 5.

Activity type and walking speed. The raw acceleration signal was downloaded and processed for two purposes. The first purpose was to develop an activity recognition algorithm using data acquired during simulated activities of daily living in the laboratory protocols. The activity recognition algorithm was then used to detect the activity types performed during the free-living protocol. Second, the raw acceleration signal was processed to determine walking speed for activities recognized as walking. The acceleration signal was segmented in nonoverlapping intervals of 5 s. This segment length was selected based on previous studies (28). Segmented data were separately filtered by two filters to create different feature sets. One feature set included accelerometer data band-pass filtered between 0.1 and 10 Hz to isolate the dynamic component due to body motion, whereas the second feature set included accelerometer data low-pass filtered at 1 Hz to isolate the static component due to gravity. The selected cut-off frequencies were based on previous research (28) and are not complementary (i.e., they are not the same cut off for both filters) due to the fact that there is no clear cut-off frequency to choose, and the two frequencies chosen were shown to be ideal in discriminating static gravitational acceleration and body motion, as shown in Fig. 2. Figure 2 shows an example of raw data, low-passed data, and band-passed data for one participant during one of the laboratory protocols. Features used for activity recognition were mean of the absolute signal, interquartile range, median, variance, main frequency peak, and low-frequency-band signal power. All accelerometer features but the median were derived from band-pass-filtered data. These features were derived and selected based on our previous work (1) using a different dataset. We report details on the mathematical formulas defined to extract accelerometer features in Table 2.

HR was extracted from RR intervals and averaged over 15-s windows. Features for the multiple linear regression model used to estimate walking speed were mean of the absolute signal, interquartile range, variance, main frequency peak, high-frequency-band signal power, and height of the participant and were also based on our previous work (2). All accelerometer features used for the walking speed models were derived from band-pass-filtered data. Coefficients for the linear regression model used to estimate walking speed are shown in Table 3.

Laboratory activities were grouped into six clusters to be used for activity classification. The six clusters were lying (lying down), sedentary (sitting, sitting and writing, standing), dynamic (cleaning the table, sweeping the floor), walking, biking, and running. Activities were derived using pattern recognition methods, in particular a Support Vector Machine (SVM). SVMs are classifiers that showed good results in classifying activities in our previous research (1, 2, 3, 4). The principle behind using pattern recognition methods and accelerometer data for activity classification is that different activity clusters (e.g., lying down, walking) result in different accelerometer patterns as collected by on-body sensors. By capturing such accelerometer patterns using the features listed in Table 2, a classifier can be trained to distinguish activity clusters with high accuracy (1–4, 28, 29). As an example, two features used for the classification of the six activity clusters are analyzed in Fig. 3. We limited the features space to two dimensions to provide a visualization that is easily human readable. Figure 3*A* shows the mean of the absolute acceleration signal, a measure representative of motion intensity. The mean of the absolute acceleration is particularly helpful in discriminating high-intensity activities (e.g., running), average-intensity activities (e.g., walking or biking), and

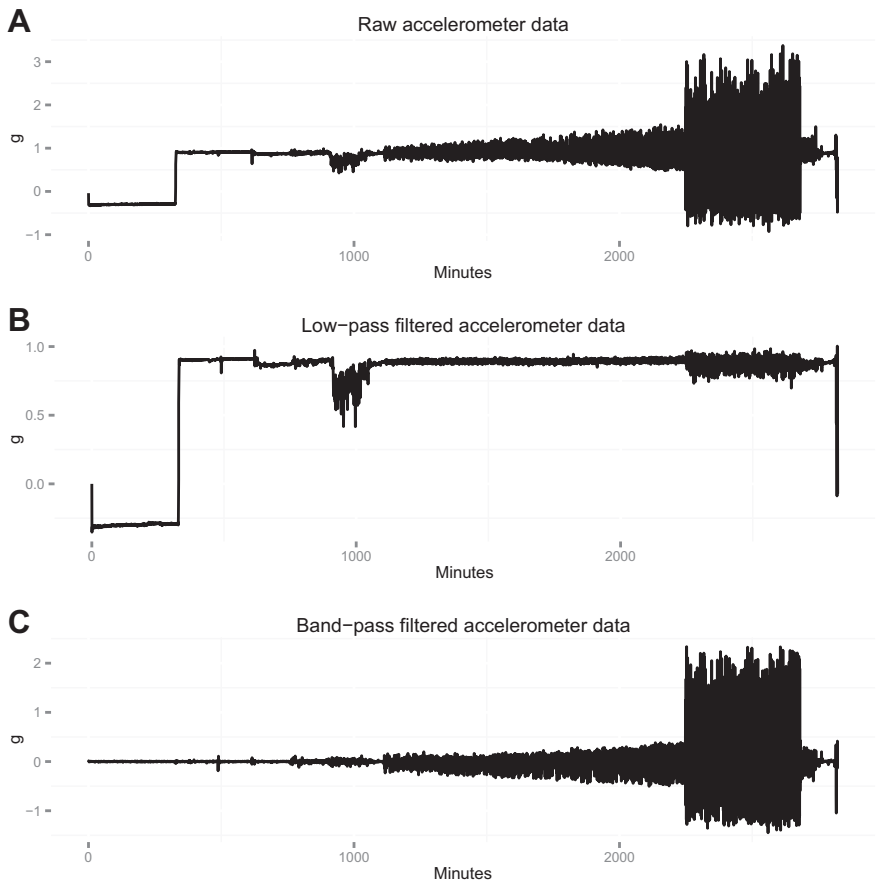

Fig. 2. Raw accelerometer data (*A*), low-pass-filtered data (*B*), and band-pass-filtered data (*C*). The gravity component is isolated when using low-pass-filtered data, as shown in *B*. This information is particularly useful to distinguish postures. Band-pass-filtered data isolate the accelerometer component due to body motion, showing increased values for higher-intensity motions. Band-pass-filtered data are particularly useful to distinguish ambulatory activities and walking speeds. Data were downsampled for visualization purposes.

low-intensity activities (e.g., lying, sedentary) as show in Fig. 3*A*. Figure 3*B* shows the median of the low-pass-filtered accelerometer *x*-axis signal, a feature representative of body posture given our sensor on-body positioning. During training, the SVM classifier takes as input multiple features (see Table 2) and determines the optimal discrimination boundary between the activity clusters, i.e., the widest separation between samples of different activity clusters (i.e., accelerometer features belonging to different activities). The distinct colored regions in Fig. 3*C* illustrate that the two shown features provide relevant information to discriminate the activity clusters. Hence, already two features are sufficient to separate most, but not all, activity clusters in this study.

Table 2. *Accelerometer features used for activity classification and walking speed estimation*

Feature Name	Computation	Description
Mean of the absolute signal	$\frac{1}{N}\sum_{t=1}^{N}\lvert a_{BPi}\rvert$	Represents motion intensity independent of the axis or orientation, similar to activity counts
Interquartile range	Q3–Q1 of a_{BP}	Represents motion intensity, can be less prone to outliers with respect to, e.g., range
Median	Middle value of the ordered a_{LP} array	Represents posture (gravitational vector)
Variance	$\frac{1}{N}\sum_{t=1}^{N}(a_i-\mu)^2$ where $\mu=\frac{1}{N}\sum_{i=1}^{N}a_i$	Represents variability in detected motion, which might be discriminative of activity type (28)
Main frequency peak	1. Apply Hamming window to reduce spectral leakage. 2. Compute FFT 3. Determine main frequency peak of the power spectrum	Provides information about the repetitiveness of motion, e.g., during walking (28)
Low-frequency band signal power	1. Apply Hamming window to reduce spectral leakage 2. Compute FFT 3. Sum signal power between 0 and 0.7 Hz (25)	Shown to be discriminative of sedentary and walking activities in previous research (25)
High-frequency band signal power	1. Apply Hamming window to reduce spectral leakage 2. Compute FFT 3. Sum signal power between 0.7 and 10 Hz (25)	Shown to be discriminative of sedentary and walking activities in previous research (25)

N, no. of samples in a 5-s window, i.e., 160 (32 samples/s); LP, low-pass-filtered data; BP, band-pass-filtered data; Q*n*, *n*th quartile; FFP, fast-Fourier transform.

Table 3. *Coefficients of the linear regression model used to estimate walking speed*

Parameter	Estimate Coefficient	*P* Value
Intercept	−1.000	$<2^{-16}$
Mean of the absolute signal	9.936	$<2^{-16}$
Variance	−2.963	$<2^{-16}$
Quartile (*x*-axis)	3.256	$<2^{-16}$
Quartile (*y*-axis)	−2.475	$<2^{-16}$
High-frequency band signal power (*x*-axis)	-5.920^{-4}	$<2^{-16}$
High-frequency band signal power (*z*-axis)	-6.320^{-4}	$<2^{-16}$
Main frequency peak (*x*-axis)	-1.323^{-1}	$<2^{-16}$
Participant height	1.439^{-2}	$<2^{-16}$

During validation all models were evaluated using leave-one-participant-out cross-validation, the coefficients shown here include all data.

The SVM trained in this paper determines decision boundaries (or separating hyperplanes) that can be used later on to classify new accelerometer feature samples into activity clusters. The decision boundaries are optimal in the sense that the algorithm determines the maximal margin between training samples and the decision boundary. Without maximizing the margin, various decision boundaries could be found. An example of a linear separation of two classes using a SVM is shown in Fig. 4. Figure 4*A* shows multiple example decision boundaries that separate the example data points, whereas Fig. 4*B* shows the separating hyperplane that maximizes the margin to the example data points, as determined by the SVM.

CRF estimation. CRF was estimated using multiple linear regression models. First, we investigated the relation between HR in specific contexts as acquired during activities of daily living simulated in laboratory settings, and $\dot{V}O_{2\,max}$. We predicted $\dot{V}O_{2\,max}$ by combining anthropometric characteristics and HR while lying down and while walking at 3.5 and 5.5 km/h. We chose lying down and walking at 3.5 and 5.5 km/h as specific contexts since lying down and walking are activities of daily living commonly performed by healthy individuals in most environments. Additionally, the average walking speeds in healthy individuals were reported in previous studies between 5 and 6 km/h (5.3 km/h in Ref. 8 and 5 ± 0.8 km/h in Ref. 19). Given the estimation error of our walking speed estimation model and the variability of free-living walking, we selected data segments with detected speed >3 and <4 km/h as segments to be considered of an average walking speed of 3.5 km/h. Similarly, we selected data

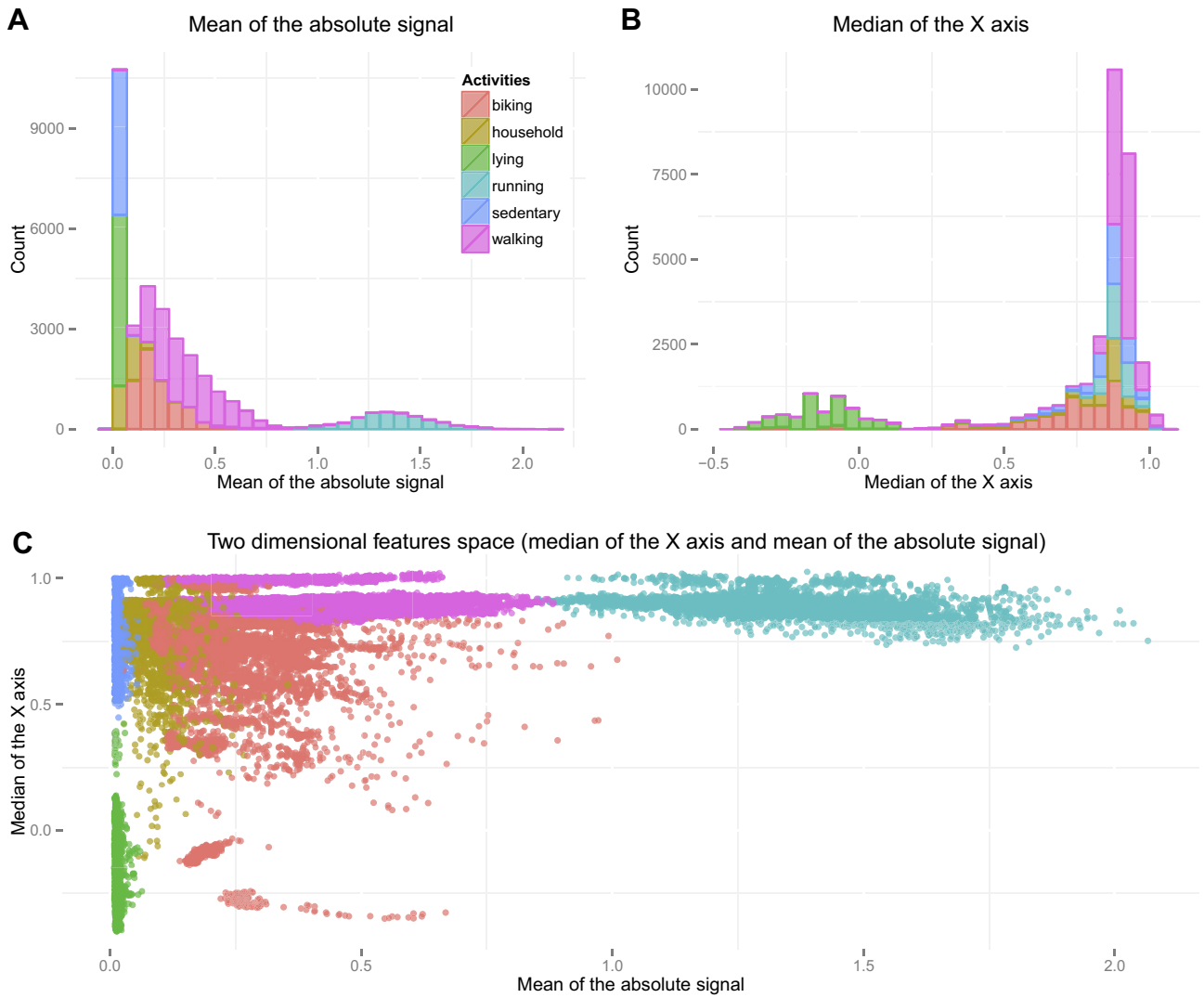

Fig. 3. Example of extracted features and multidimensional features space used for activity classification of the six activity clusters included in this study. *A* and *B*: histograms of two accelerometer features (mean of the absolute signal and median of the *x*-axis). *C*: two-dimensional features space showing clear separations between most activity clusters.

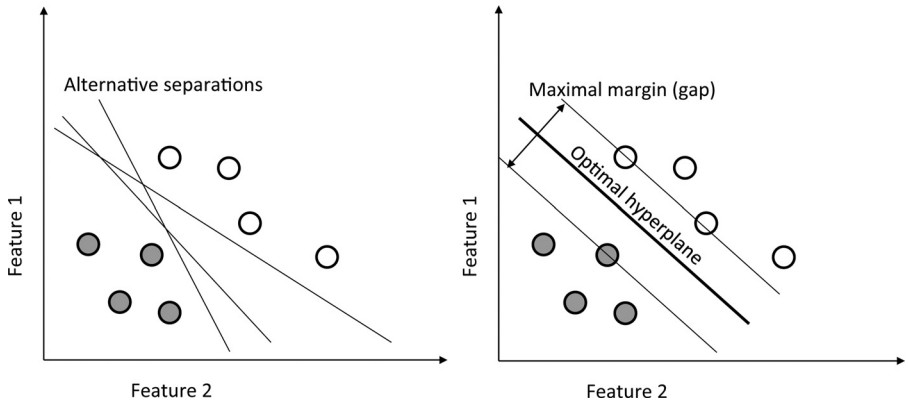

Fig. 4. Example of linear decision boundaries to classify two classes. Example data points of the classes are illustrated in different gray tones. *Left*, different example decision boundaries; *right*, optimal hyperplane obtained by maximizing the margin between the decision boundary and the closest data points of each class. Samples on the maximal margin lines are called support vectors.

segments with detected speed >5 and <6 km/h as segments to be considered of an average walking speed of 5.5 km/h.

Next, we analyzed the relation between context-specific HR during activities of daily living simulated in laboratory settings, and context-specific HR during the same activities as detected by our activity recognition and walking speed models, in free living. The analysis of the relation between context-specific HR in laboratory settings and free living consisted of computing the correlation coefficient and relative differences between HR in laboratory settings and free living. This analysis is merely to provide some perspective on context-specific HR with respect to laboratory measurements. However, free-living regression models are built and evaluated using free-living data only.

Finally, we predicted $\dot{V}O_{2\,max}$ by combining anthropometric characteristics and HR while lying down and while walking at 3.5 and 5.5 km/h as determined from free-living data to evaluate the ability of the context-specific HR detected using pattern recognition methods to estimate CRF.

Statistics

Activity recognition and walking speed estimation models were derived using laboratory data and evaluated using leave-one-participant-out cross-validation. The same training set, consisting of data from all participants but one, was used to build feature selection, activity recognition, and walking speed estimation and CRF estimation models. The remaining data were used for validation. The procedure was repeated for each participant, and results were averaged. Performance of the activity recognition models was evaluated using the class-normalized accuracy, using laboratory recordings. Results for walking speed estimation were reported in terms of root mean-square error (RMSE), where the outcome variable was speed in kilometers per hour. The relations between HR and CRF were reported using Pearson's correlation coefficient (r) for both activities simulated in laboratory settings and free-living data. The relation between context-specific HR during activities of daily living simulated in laboratory settings and in free living as detected by pattern recognition methods was reported using Pearson's correlation coefficient (r) and the mean and SD of the difference between context-specific HR in laboratory settings and in free living. Results for CRF estimation models were reported in terms of explained variance (R^2). The Bland-Altman plot was used to determine the agreement between measured and predicted CRF. Finally, subject-independent evaluation for CRF estimation models was also performed, using leave-one-participant-out cross-validation. Regression models including different HR parameters (e.g., HR while lying down or HR while walking at different speeds) were compared using the likelihood ratio. More specifically, we compared two models, the first one including anthropometrics and HR while lying down and a second one including anthropometrics, HR while lying down, and HR while walking. We compared likelihood ratios for both laboratory recordings and free-living data. We reported results for subject-independent CRF estimation in terms of RMSE, where the outcome variable was $\dot{V}O_{2\,max}$ in milliliters per minute as measured in laboratory conditions. Paired t-tests were used to compare results. Significance was set at $\alpha < 0.05$.

RESULTS

Descriptive Statistics

The dataset considered for this work contained 491 days of data collected from 51 participants in free living, thus ~10 days/participant, including accelerometer and ECG data. Eighty-three hours of laboratory recordings, including reference $\dot{V}O_2$, $\dot{V}CO_2$, acceleration, ECG, and $\dot{V}O_{2\,max}$, were collected for model building and evaluation. Laboratory measurements were discarded for two participants where we observed measurement errors such as unusable ECG data due to excessive noise or bad lead attachment. Anthropometric characteristics and CRF level for the participants are reported in Table 1. Figure 5 shows an exemplary output of the walking speed and activity recognition models for one participant during 24 h of free-living recordings. Context-specific HR as identified using activity recognition and walking speed models in free living is also shown in Fig. 5.

CRF Estimation from Context-Specific Submaximal HR during Simulated Activities of Daily Living

HR during activities of daily living simulated in laboratory settings was 66.2 ± 12.3 beats/min for lying, 91.0 ± 15.3 beats/min for walking at 3.5 km/h, and 107.8 ± 17.7 beats/min for walking at 5.5 km/h. Pearson's correlation between context-specific submaximal HR as measured during activities of daily living simulated in laboratory settings and CRF was −0.43 for lying down, −0.47 for walking at 3.5 km/h, and −0.51 for walking at 5.5 km/h, thus confirming the hypothesis that submaximal HR is inversely related to CRF. Explained variance (adjusted R^2) for multiple regression models including sex, body weight, and age as predictors of CRF was 0.64. Adjusted R^2 increased when including context-specific HR and was 0.69 for lying, 0.72 for walking at 3.5 km/h, and 0.74 for walking at 5.5 km/h, thus confirming that activities of higher submaximal intensities explain more of the variance in the model. Results are reported in Table 4 while Fig. 6 shows scatterplots of reference against fitted values and Bland-Altman plots. When including more advanced anthropometrics, such as fat-free mass instead of body weight, R^2 was 0.73 when no HR was used among the

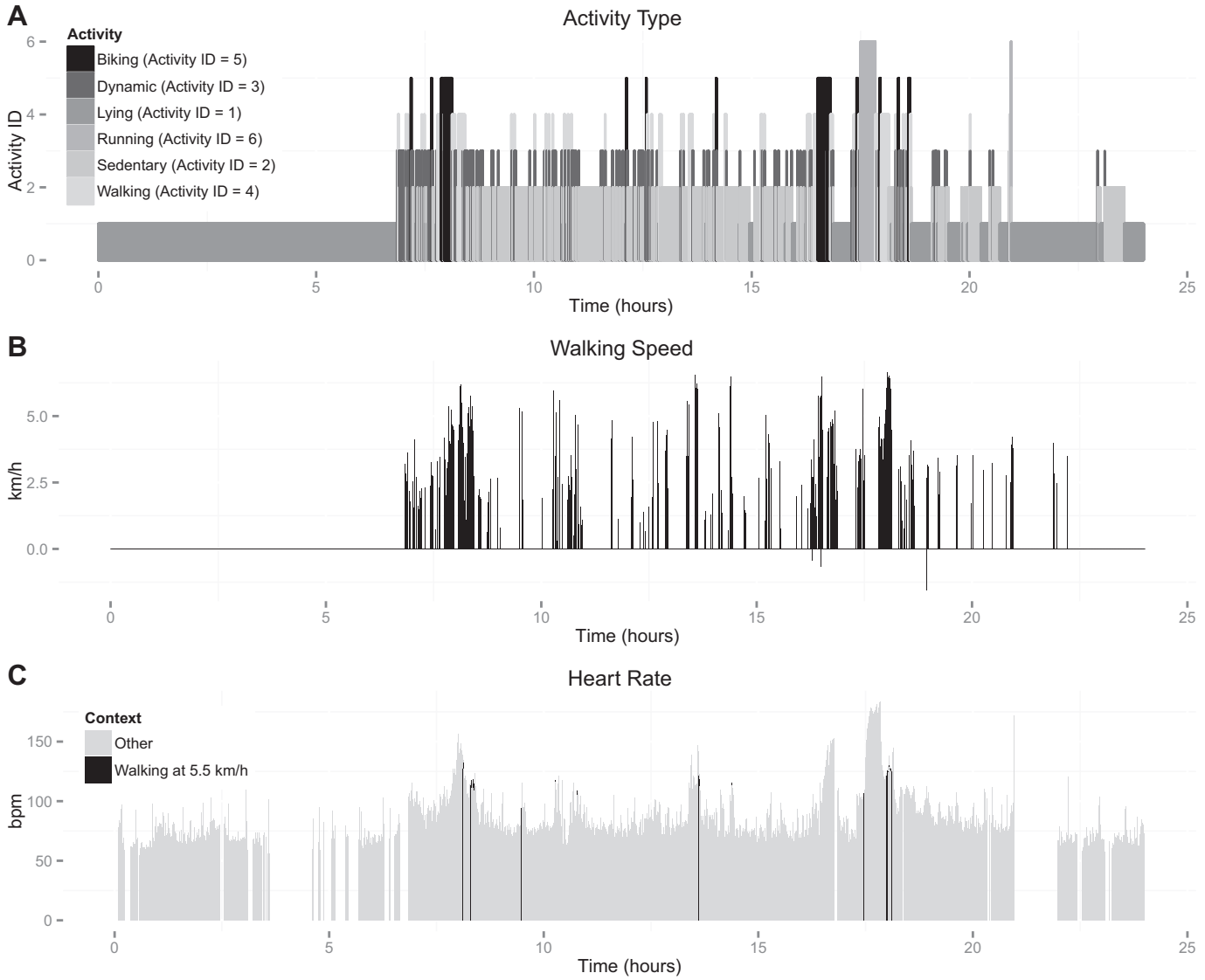

Fig. 5. Exemplary output of the models used to contextualize HR in free living in this work, for one participant. *A*: recognized activity types. Commuting by bike, training (running), sleeping, and a mostly sedentary job during waking hours can be easily identified from this plot. *B*: estimated walking speeds when the activity type algorithm identifies the walking activity. *C*: HR and contextualized HR. Contextualized HR, i.e., in this example the HR while walking at 5.5 km/h, is highlighted in black.

predictors, 0.74 for lying, 0.76 for walking at 3.5 km/h, and 0.78 for walking at 5.5 km/h. We computed the likelihood ratio between regression models including anthropometrics data and HR while lying down with respect to regression models including anthropometrics data and HR while walking at 3.5 and 5.5 km/h. The likelihood ratio showed that, for both walking speeds, including HR while walking significantly improved the model fit ($P = 0.044$ when including

Table 4. *Multiple linear regression models for $\dot{V}O_{2max}$ estimation from activities of daily living simulated in laboratory settings*

Model Description	Predictors	R^2
Anthropometric characteristics only	Intercept (1431.686, $P = 0.00322$), body wt (18.510, $P = 0.00645$), age (−1.888, $P = 0.84595$), sex (798.561, $P = 7.48^{-7}$)	0.64
Context-specific HR	Intercept (2849.294, $P = 3.86e\text{-}05$), HR while lying down in laboratory settings (−14.268, $P = 0.00344$), body wt (13.862, $P = 0.02862$), age (−8.170, $P = 0.37335$), sex (803.669, $P = 1.20^{-7}$)	0.69
	Intercept (3183.644, $P = 5.67e\text{-}06$), HR while walking at 3.5 km/h in laboratory settings (−13.636, $P = 0.000496$), body wt (14.921, $P = 0.013229$), age (−12.046, $P = 0.183661$), sex (777.430, $P = 1.03^{-7}$)	0.72
	Intercept (3367.865, $P = 9.65^{-7}$, HR while walking at 5.5 km/h in laboratory settings (−13.044, $P = 8.21^{-5}$), body wt (15.234, $P = 0.00856$), age (−13.222, $P = 0.13055$), sex (754.772, $P = 8.95^{-8}$)	0.74

For each predictor, detailed information (model coefficient, *P* value) is indicated. Data are for 49 subjects. HR, reart rate.

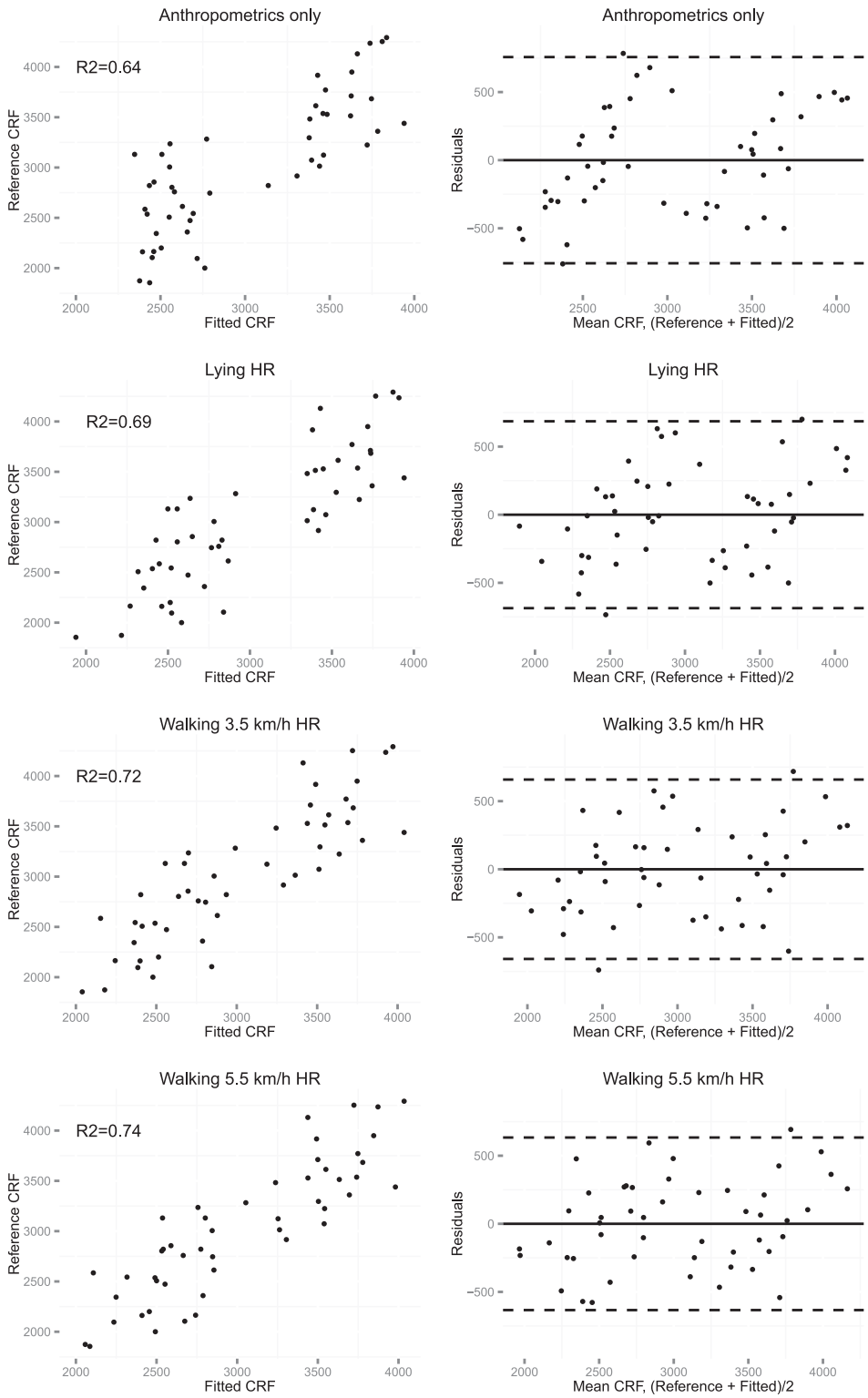

Fig. 6. Accuracy of the prediction models for CRF estimation. Regression plots and Bland-Altman plots are shown for models using as predictors anthropometrics and context-specific HR during activities of daily living simulated in laboratory conditions. R^2 is also reported.

the HR while walking at 3.5 km/h and $P = 0.0048$ when including the HR while walking at 5.5 km/h).

Context Recognition, Activity Type, and Walking Speed

Laboratory recordings with reference activity type were used to determine accuracy of the models used in free living. Accuracy of the SVM activity recognition classifier was 94.1%. More specifically, the accuracy was 96.4% for lying, 95.6% for sedentary activities, 83.3% for dynamic, 98.2% for walking, 91.4% for biking, and 99.7% for running. The confusion matrix for the subject-independent results of the activity recognition model is shown in Table 5. The explained variance for the walking speed model was 0.85 (R^2). Walking speed estimation RMSE for subject-independent analysis was 0.37 km/h across all speeds.

Table 5. *Confusion matrix showing the normalized performance of the activity recognition model*

		Classification Results					
True Activities		Lying	Sedentary	Dynamic	Walking	Biking	Running
	Lying	96	4	0	0	0	0
	Sedentary	0	96	4	0	0	0
	Dynamic	0	10	83	0	7	0
	Walking	0	0	0	98	2	0
	Biking	0	1	4	5	90	0
	Running	0	0	0	0	0	100

Units are %.

Activities in free living over the complete dataset were recognized as follows: 44.4% lying, 36.4% sedentary, 9.5% dynamic, 5.4% walking, 3.8% biking, and 0.4% running. Average walking speed was 3.6 ± 1.5 km/h. Participants spent on average 77.7 min/day walking, 11.9 min of which were at 3.5 km/h and 11.6 min of which were at 5.5 km/h.

Relation between Context-Specific Submaximal HR during Activities of Daily Living Simulated in Laboratory Settings and in Free Living

Pearson's correlation between context-specific submaximal HR measured during activities of daily living simulated in laboratory settings and in free living as detected by pattern recognition methods was 0.71 for lying down, 0.71 for walking at 3.5 km/h, and 0.75 for walking at 5.5 km/h. Mean difference between context-specific HR in laboratory settings and free living was 2.9 ± 8.7 for lying (mean HR while lying down was 63.2 beats/min in free living and 66.2 beats/min in laboratory settings), 8.7 ± 11.2 for walking at 3.5 km/h (mean HR while walking at 3.5 km/h was 99.9 beats/min in free living and 91.0 beats/min in laboratory settings), and −2.7 ± 11.5 for walking at 5.5 km/h (mean HR while walking at 5.5 km/h was 106.3 beats/min in free living and 107.8 beats/min in laboratory settings). Thus, all differences were below 10%. Histograms of the differences and scatterplots of context-specific HR in laboratory settings and free living are shown in Fig. 7.

CRF Estimation from Context-Specific Submaximal HR in Free Living

HR during specific contexts in free living was 63.2 ± 9.3 beats/min for lying, 99.9 ± 11.6 beats/min for walking at 3.5 km/h, and 106.3 ± 11.8 beats/min for walking at 5.5 km/h.

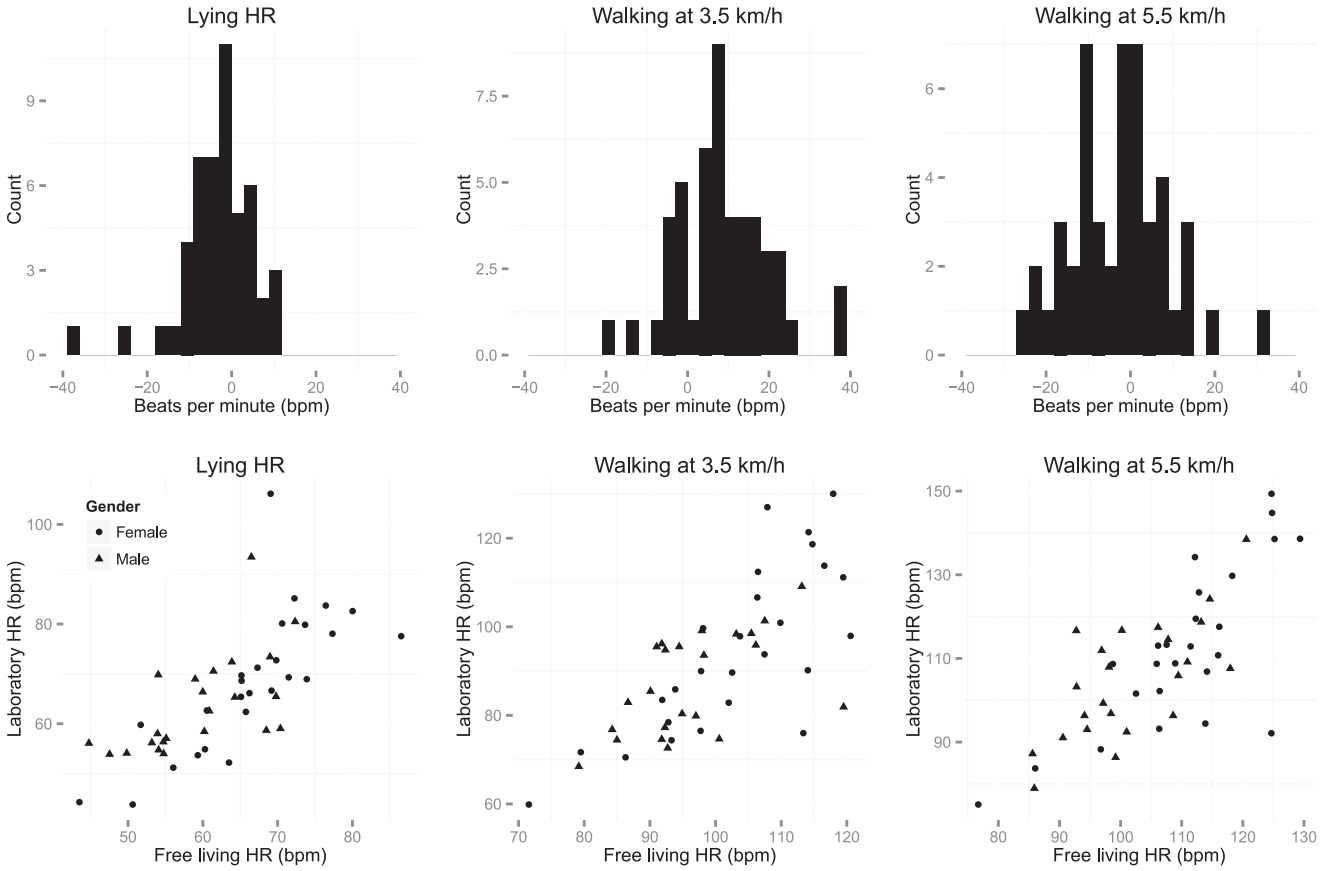

Fig. 7. *Top*, histograms of differences between context-specific HR in laboratory settings and free living. *Bottom*, scatterplots showing the relation between context-specific HR in laboratory settings and free living.

Table 6. *Multiple linear regression models for $\dot{V}O_{2max}$ estimation from free-living data*

Model Description	Predictors	R^2
Anthropometric characteristics only	Intercept (1403.603, $P = 0.00326$), body wt (19.531, $P = 0.00355$), age (−3.184, $P = 0.73931$), sex (803.869, $P = 4.59^{-7}$)	0.65
Context-specific HR	Intercept (2914.307, $P = 6.31^{-6}$), HR while lying down in free living (−22.118, $P = 0.000554$), body wt (21.150, $P = 0.000511$), age (−9.027, $P = 0.298184$), sex (634.875, $P = 1.32^{-5}$)	0.73
	Intercept (4175.338, $P = 2.19^{-6}$), HR while walking at 3.5 km/h in free living (−20.798, $P = 0.000136$), body wt (16.106, $P = 0.005611$), age (−20.240, $P = 0.032176$), sex (738.579, $P = 1.55^{-7}$)	0.74
	Intercept (4647.138, $P = 1.03^{-7}$), HR while walking at 5.5 km/h in free living (−23.884, $P = 7.03^{-6}$), body wt (16.801, $P = 0.0022$), age (−21.322, $P = 0.0156$), sex (668.687, $P = 5.02^{-7}$)	0.77

For each predictor, detailed information (model coefficient, P value) is indicated. Data are for 51 subjects.

Pearson's correlation between context-specific submaximal HR as measured in free living and CRF was −0.54 for lying down, −0.52 for walking at 3.5 km/h, and −0.60 for walking at 5.5 km/h, thus confirming the hypothesis that submaximal HR is inversely related to CRF. Adjusted R^2 increased from the case where no HR was included ($R^2 = 0.65$), when including context-specific HR. More specifically R^2 was 0.73 for lying, 0.74 for walking at 3.5 km/h, and 0.77 for walking at 5.5 km/h, thus confirming that activities of higher submaximal intensities explain more of the variance in the model, even when carried out in free living. Results for all models are reported in Table 6, and Bland-Altman plots for all models are shown in Fig. 8. When including more advanced anthropometrics, such as fat-free mass instead of body weight, R^2 was 0.73 when no HR was used among the predictors, 0.77 for lying, and 0.80 for walking at 3.5 and 5.5 km/h. We computed the likelihood ratio between regression models including anthropometrics data and HR while lying down with respect to regression models including anthropometrics data and HR while walking at 3.5 and 5.5 km/h. The likelihood ratio showed that, for both walking speeds, including HR while walking significantly improved the model fit ($P = 0.0047$ when including the HR while walking at 3.5 km/h and $P = 0.00027$ when including the HR while walking at 5.5 km/h).

Cross-Validation of $\dot{V}O_{2\,max}$ Estimates

$\dot{V}O_{2\,max}$ estimation models derived from free-living data were cross-validated using the leave-one-out technique. Results are reported in Tables 7 and 8. Cross-validation of $\dot{V}O_{2\,max}$ estimates using as predictors context-specific HR as measured during activities of daily living simulated in laboratory settings, RMSE for the model including anthropometric characteristics only as predictors was 358.3 ml/min (R^2 was 0.66). RMSE was reduced when including HR in specific contexts among the predictors, with RMSE = 314.3 ml/min ($R^2 = 0.73$) for lying down, RMSE = 310.0 ml/min ($R^2 = 0.75$) for walking at 3.5 km/h, and RMSE = 284.7 ml/min ($R^2 = 0.78$) for walking at 5.5 km/h as specific contexts. Thus, RMSE was reduced up to 21% when including context-specific HR among the predictors. Cross-validation of $\dot{V}O_{2\,max}$ estimates using as predictors context-specific HR as derived by pattern recognition methods in free living, RMSE for the model including anthropometric characteristics only as predictors was 354.7 ml/min (R^2 was 0.67). RMSE was reduced when including HR in specific contexts among the predictors, with RMSE = 309.4 ml/min ($R^2 = 0.75$) for lying down, RMSE = 305.91 ml/min ($R^2 = 0.76$) for walking at 3.5 km/h, and RMSE = 281.0 ml/min ($R^2 = 0.79$) for walking at 5.5 km/h as specific free-living contexts. Thus, RMSE was also reduced up to 21% when including context-specific HR as determined from pattern recognition methods among the predictors.

DISCUSSION

In this work, we proposed a method to estimate $\dot{V}O_{2\,max}$ in free living without the need for laboratory tests or specific protocols. While many methods have been developed to estimate $\dot{V}O_{2\,max}$ using data collected under supervised laboratory conditions or following strict protocols, limited work tried to estimate CRF using wearable sensors and data collected under unsupervised settings in free living (9, 22). We adopted pattern recognition techniques to determine specific contexts, e.g., low-intensity activities of daily living such as lying down and walking at predefined speeds, to contextualize submaximal HR without the need for a strict exercise protocol. We first validated the effectiveness of submaximal context-specific HR as a predictor of $\dot{V}O_{2\,max}$ during activities of daily living simulated in laboratory settings. Next, we analyzed the correlation and relative differences between context-specific HR during activities simulated in the laboratory and context-specific HR as detected by pattern recognition methods deployed in free living. Finally, we used context-specific HR in free living to estimate CRF. Our results showed that $\dot{V}O_{2\,max}$ estimation using as predictors context-specific HR in free living provides accuracy comparable with laboratory-derived models. In particular, RMSE for $\dot{V}O_{2\,max}$ estimation could be reduced up to 21% compared with anthropometric characteristics only by using as predictors HR in specific contexts as determined by pattern recognition methods in free living.

Context-Specific HR during Activities of Daily Living Simulated in Laboratory Settings

The main assumption behind this study was that submaximal HR is inversely related to $\dot{V}O_{2\,max}$ and that the correlation is higher during submaximal activities of higher intensity. Our laboratory recordings confirm this assumption. Pearson's correlation between context-specific HR and $\dot{V}O_{2\,max}$ went from −0.43 to −0.51 for lying and walking activities. Multiple regression models showed higher explained variance (R^2 between 0.64 and 0.74) when including context-specific HR. Increasing activity intensity, i.e., from lying to slow walking (3.5 km/h) to faster walking (5.5 km/h), further improved R^2. Finally, the likelihood ratio showed that model fit improved

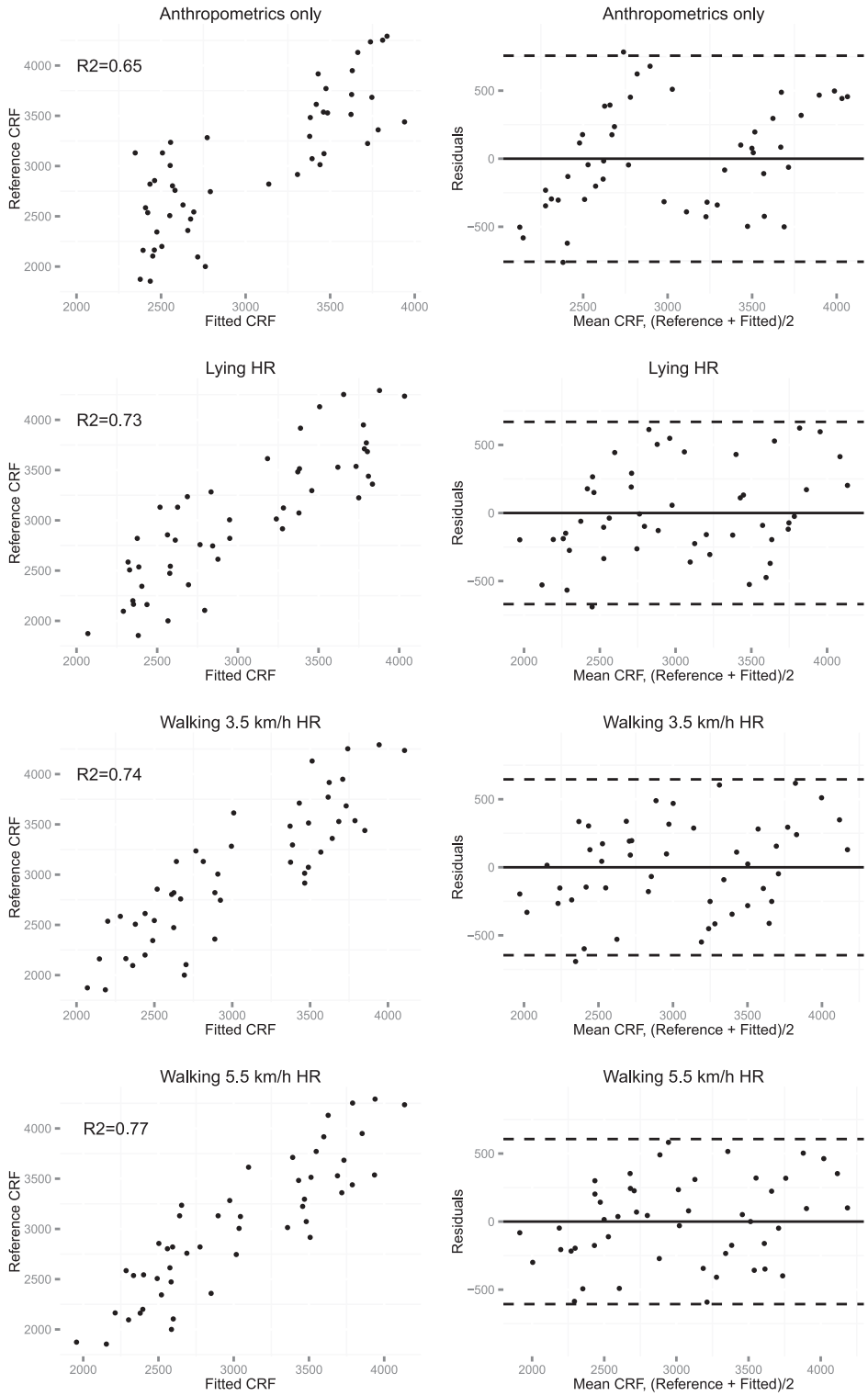

Fig. 8. Accuracy of the prediction models for CRF estimation. Regression plots and Bland-Altman plots are shown for models using as predictors anthropometrics and context-specific HR in free living. R^2 is also reported.

significantly when including in the regression models not only HR while lying down but also HR while walking at different speeds. These results are in agreement with a significant body of literature relying on submaximal HR for $\dot{V}O_{2\,max}$ estimation during more intense activities, such as biking or running, compared with the low-intensity activities used in this study (30).

Context Recognition in Free Living

We deployed activity recognition and walking speed estimation algorithms in free living to contextualize submaximal HR without the need for strict exercise protocols or laboratory tests. Our activity recognition model showed high accuracy in detecting lying and walking activities (96.4–98.2%) given the

Table 7. *Cross-validation of multiple linear regression models for $\dot{V}O_{2max}$ estimation using as predictors context-specific HR as measured during activities of daily living simulated in laboratory settings*

Model Description	Predictors	RMSE, ml/min	R^2
Anthropometric characteristics only	Body wt, age, sex	358.3	0.66
Context-specific HR	HR while lying down in laboratory settings, body wt, age, sex	314.3	0.73
	HR while walking at 3.5 km/h in laboratory settings, body wt, age, sex	310.0	0.75
	HR while walking at 5.5 km/h in laboratory settings, body wt, age, sex	284.7	0.78

RMSE, root mean-square error.

characteristic accelerometer fingerprints of such activities, characterized either by different accelerometer orientation with respect to other activities or very specific repetitive movements. The activities chosen as free-living contexts were lying down and walking for the following reasons. First, those are common activities performed by healthy individuals in most environments. Second, the inverse relation between HR at rest or sleeping HR and CRF was already shown in previous research, highlighting how this parameter can be valuable for $\dot{V}O_{2\,max}$ estimation (17, 30). Finally, walking activities can be discriminated in intensity, by detecting walking speed, using simply an accelerometer. This is an important factor when trying to detect specific context in free living, since detecting only activity type, if the activity can be carried out at different intensities, would not be sufficient to determine the same context for each individual. However, walking is an activity that can be accurately quantified in terms of both type (i.e., walking) and intensity (i.e., speed). The proposed activities are low intensity and were performed daily by the participants involved in our study, as shown by the analysis of free-living data. Our study population spent on average 44.4% of the free-living time lying down and 5.4% of the free-living time walking. Of the time spent walking, 11.9 min daily were spent at 3.5 km/h while 11.6 min daily were spent at 5.5 km/h, the two speeds used by our models to contextualize HR. Considering that many fitness tests require protocols shorter than 11 min (e.g., the common 6-min walking test), we believe a total of 10 min daily is a sufficient amount of data for prediction of $\dot{V}O_{2\,max}$, at least in the population of healthy adults considered in this study. We could evaluate activity recognition and walking speed models only under laboratory conditions where reference was present. Among the recognized activities, the dynamic activity cluster was recognized with accuracy below average (see confusion matrix). We interpret that activities with high variability in movement and execution between participants and using a single chest-worn sensor resulted in higher classifier confusion. However, the high accuracy of walking speed estimation models and activity recognition for walking provide confidence for the free-living detection of activities used to contextualize HR. Additionally, from the cross-validation analysis results we can see how subject-independent models built using activities of daily living simulated in laboratory settings (RMSE were 314.3, 310.0, and 284.7 ml/min for lying, walking at 3.5 km/h, and walking at 5.5 km/h, respectively) are similar to RMSE results obtained contextualizing HR using pattern recognition methods in free living (309.4, 305.9, and 281.0 ml/min for lying, walking at 3.5 km/h, and walking at 5.5 km/h, respectively). These results can serve as indirect validation of the accuracy of activity recognition and walking speed estimation in properly detecting the relevant contexts in free living.

Context-Specific HR in Free Living

Context-specific HR in free living showed relations with $\dot{V}O_{2\,max}$ similar to what we reported in laboratory settings. The inverse relation between HR at a certain workload and $\dot{V}O_{2\,max}$ is the key principle behind laboratory-based submaximal CRF tests, and this relation proved to be valid not only in laboratory settings but also in free living as well. The correlation between HR while lying down in free living and $\dot{V}O_{2\,max}$ was −0.54, and it was increased up to −0.60 when the HR while walking at 5.5 km/h in free living was used, highlighting how activities of higher intensity result a stronger link between submaximal HR and $\dot{V}O_{2\,max}$. Explained variance also increased between 0.65 when anthropometrics characteristics only were used to estimate $\dot{V}O_{2\,max}$ and 0.77 when using context-specific HR. Finally, the likelihood ratio showed that model fit improved significantly when including in the regression models not only HR while lying down but also HR while walking at different speeds. We also analyzed the relation between HR during the same activities carried out in laboratory settings and free living. We expected differences in HR due to the different settings, e.g., walking in free living might include carrying weights, walking on inclined surfaces, or other factors that might raise HR. On the other hand, lying down in laboratory settings might be more stressful than sleeping, therefore lowering HR with respect to laboratory conditions. Additionally, a single laboratory measurement might be affected by factors such as the previous day's physical activity, while free-living recordings averaged over multiple days might provide more stable representations of a participant's physiology. On the

Table 8. *Cross-validation of multiple linear regression models for $\dot{V}O_{2max}$ estimation using as predictors context-specific HR as detected by pattern recognition methods in free living*

Model Description	Predictors	RMSE, ml/min	R^2
Anthropometric characteristics only	Body wt, age, sex	354.7	0.67
Context-specific HR	HR while lying down in free living, body wt, age, sex	309.4	0.75
	HR while walking at 3.5 km/h in free living, body wt, age, sex	305.9	0.76
	HR while walking at 5.5 km/h in free living, body wt, age, sex	281.0	0.79

other hand, free-living data might include more bouts of fragmented walking and therefore HR might not always reach steady state. Thus, the relation between HR during activities simulated in laboratory conditions and between HR and free-living activities is most likely different, and models deployed in free living should be developed using free-living data, as proposed by our methodology. However, analyzing the relation between laboratory and free-living HR in the same contexts can be useful to determine to what extent laboratory recordings can be reproduced in free living as well as the ability of pattern recognition methods to detect differences between contexts such as lying down or walking at different speeds in unsupervised free-living conditions. The relatively high correlation between laboratory and free-living HR (0.71–0.75), as well as similar mean values and consistent differences between conditions (i.e., higher HR for walking at higher speed, or higher intensity, in our case HR for laboratory activities and free living was 66.2 and 63.2 beats/min for lying, 91.0 and 99.9 for walking at 3.5 km/h, and 107.8 and 106.3 for walking at 5.5 km/h), are all promising results that free-living data can be used as a reliable substitute of laboratory recordings for context-specific submaximal HR.

Fat-Free Mass

Analysis of $\dot{V}O_{2\,max}$ estimation including fat-free mass instead of body weight among the predictors resulted in higher accuracy as expected and previously shown in the literature (22). In particular, R^2 was increased between 0.74 and 0.78 for laboratory-based measurements and between 0.77 and 0.80 for context-specific HR determined in free living. However, because the aim of our work is to provide $\dot{V}O_{2\,max}$ estimation outside of the laboratory environment, we focus on simple anthropometrics only (i.e., body weight, age, and sex) in the remaining of our discussion.

Cross-Validation of $\dot{V}O_{2\,max}$ Estimates

We also performed cross-validation using subject-independent models for $\dot{V}O_{2\,max}$ estimation, since our aim was to validate the proposed methods using state of the art techniques able to validate the model on unseen data. Results for cross-validation were consistent with what was shown before. Our results confirm that, when estimating CRF, the individual's anthropometric characteristics are not sufficient to provide an accurate estimate. Differences in CRF among participants with similar body size (e.g., similar body weight and height) are not distinguishable if no physiological data are used in the models. Thus, the lower RMSE shown by $\dot{V}O_{2\,max}$ estimation models including HR as predictor shows the ability of submaximal context-specific HR to discriminate between such participants with similar anthropometric characteristics and further reduce $\dot{V}O_{2\,max}$ estimation error. As expected, contextualizing HR using more intense activities, such as walking at 5.5 km/h instead of lying, provides better results. It is interesting to note that subject-independent analysis RMSE was reduced consistently between models using anthropometrics only and context-specific HR (for any activity) both in laboratory settings and free living. However, increasing the intensity of the specific context analyzed, e.g., from lying down to walking at 3.5 km/h to walking at 5.5 km/h, did not consistently reduce RMSE. RMSE for models including HR while lying down and slow walking (i.e., walking at 3.5 km/h) were similar, highlighting that the physiological responses to exercise we are interested in monitoring might require a certain level of intensity for the model to benefit beyond what can be already achieved using lying HR as a predictor. These findings are valid both in laboratory settings using HR during simulated activities of daily living and in free living using HR as detected by pattern recognition methods.

Comparison with Prior Work

Little work was reported in the literature on protocol-free $\dot{V}O_{2\,max}$ estimation. Previous studies aiming at estimating $\dot{V}O_{2\,max}$ in free-living conditions were either limited to using physical activity-related parameters, such as steps, as proposed by Cao et al. (9), HR normalized by activity intensity, as proposed by Plasqui and Westerterp (22), or requiring intense exercise such as running (32). Results for $\dot{V}O_{2\,max}$ estimation reported in terms of R^2 or RMSE cannot be easily compared between studies due to the dependency of these parameters on the study participant characteristics, e.g., body weight and $\dot{V}O_{2\,max}$ levels. However, we report in this section R^2 results as typically reported by other studies to put ours in perspective with current state of the art in $\dot{V}O_{2\,max}$ estimation. For some studies, e.g., Ref. 21, participants had similar characteristics to our study, and therefore comparisons can be meaningful. We reported R^2 of 0.79 for our subject-independent analysis. Results reported by Plasqui and Westerterp on a cross-validation sample for their method showed that using as predictor HR divided by activity counts, a measure of motion intensity, $\dot{V}O_{2\,max}$ could be predicted with $R^2 = 0.72$. The populations in the two studies are comparable, and therefore further contextualizing HR in free living (i.e., using as predictor HR while walking at a certain speed) seems beneficial. Other protocols involving more intense activities, such as running, did not provide better results. For example, by combining the ratio of inverse foot-ground contact time and HR during steady-state running, Weyand et al. (32) reported $R^2 = 0.74$ in the experimental group and $R^2 = 0.67$ in the cross-validation group.

By using context-specific HR in free living as a predictor, we obtained results comparable to or better than previous free-living studies and also comparable to what was reported using similar metrics in laboratory settings or while performing strict protocols (25). For example, 92 different $\dot{V}O_{2\,max}$ protocols were reviewed in a recent analysis by Sartor et al. (27). Additionally to the free-living studies discussed here, the authors suggested that many other submaximal tests could be performed in free living without laboratory infrastructure. However, most of these tests require intense activities and strict protocols, e.g., the most commonly used 2-mile run [$R^2 = 0.81$; Mello et al. (18)], Canadian aerobic fitness test [$R^2 = 0.82$; Jetté et al. (13)], or YMCA [$R^2 = 0.56$; Santo and Golding (26)]. The accuracy of the best-performing tests is comparable to our free-living estimation. However, the approach proposed in this work does not require intense activities and is therefore suitable on a wider population. Additionally, the proposed approach does not require a specific test, and therefore $\dot{V}O_{2\,max}$ could be continuously assessed longitudinally over time and not only reassessed when the test is performed. The effectiveness of context-specific HR as derived in free living with respect to laboratory-based protocols was

also validated in our own analysis showing comparable RMSE and R^2 when including laboratory-derived HR or free-living HR.

Other studies investigate the relation between easily accessible measures such as HR or HR variability at rest and $\dot{V}O_{2\,max}$ (11). However, these studies typically reported low levels of accuracy [$R^2 = 0.29$; Esco et al. (11)], showing that single measurements or "spot" measurements of physiological parameters and limited levels of context are insufficient for a reliable $\dot{V}O_{2\,max}$ estimate. A possible explanation for the better performance of the proposed approach compared with both single spot checks (11) and more intense protocols that can be carried out in free living is that, by contextualizing HR over multiple days, our proposed approach is less prone to the day-to-day variability typical of physiological measurements.

The clear advantage of the current approach is the ability to provide estimates during normal activities of daily living, as carried out by individuals. We validated our models independently on the participant, using cross-validation and the leave-one-out technique. Additionally, for all of our models, we also computed results using as predictor body weight instead of fat-free mass, thus providing estimates from easily accessible measures that can be acquired without complex and expensive laboratory infrastructure. Our results are extendable to new participants without the need of retraining the models or other laboratory protocols. The current implementation could be directly deployed to new studies in free-living conditions.

Limitations and Future Work

A limitation of this study is the validation on healthy adults only, with similar lifestyles in a Dutch setting. Future work should investigate if the proposed CRF estimation model is suitable for other groups such as the obese and persons affected by chronic disease, and if the proposed activity recognition system or other activity recognition systems trained to recognize only the relevant activities to contextualize HR (e.g., lying and walking) can be suitable for these populations. In nonhealthy populations changes in CRF could provide an additional marker of disease progression. Additionally, future work should address the ability of the proposed method not only to estimate CRF for an individual but to track changes in CRF over time, e.g., by means of a physical activity intervention. In this study, we assumed $\dot{V}O_{2\,max}$ to remain constant over a period of 2 wk, since participants were not implementing changes to their lifestyle, and typical interventions to modify $\dot{V}O_{2\,max}$ are of much longer duration [e.g., 3 mo to 1 yr (14)]. Finally, in this study we used a wearable sensor prototype (the ECG Necklace) to collect data. The ECG Necklace provided raw accelerometer and ECG data streams that were processed to determine activity type, and HR. While the heart beat detection and activity recognition algorithms are not detailed in this paper, these basic processing components are replaceable and well known in the literature (1, 24, 28), and the novelty of our contribution is in the methodology of using the components to contextualize HR in free living so that we could validate our hypothesis of estimating $\dot{V}O_{2\,max}$ using only free-living data. Thus, this study can be completely replicated by using off-the-shelf sensors for accelerometer and HR recordings instead of the ECG Necklace prototype, since many wearable sensors able to detect activities and HR are available on the market today. Especially, heart activity sensors today mostly provide HR data and not ECG, simplifying the analysis procedure.

CRF is a strong and independent predictor of all-cause and cardiovascular mortality. When evaluating the suitability and practical applicability of a new test, many parameters should be accounted for. The cost, convenience, and infrastructure required are current barriers to widespread $\dot{V}O_{2\,max}$ measurements, despite the well-known relevance in healthcare. The proposed CRF estimation model is applicable to a wide population, since it does not require intense physical exercise and requires accelerometer and HR data only. Such measures are becoming more and more widespread due to mainstream availability of wearable technology, including combined accelerometer and HR monitors. Similarly, the processing capabilities of modern mobile phones are sufficient for practical deployment of machine learning methods (4).

In conclusion, this work showed that contextualized HR in free living can be used to provide $\dot{V}O_{2\,max}$ estimation with accuracy comparable to other methods relying on submaximal HR measured in laboratory settings. This is the first study using pattern recognition methods to automatically contextualize HR in free living and predict CRF. We showed that considering context-specific HR provides better CRF estimates, and including context-specific HR at higher intensities (e.g., while walking) further reduces estimation error. Additionally, we show increased accuracy depending on activity intensity. When including HR while walking in the estimation model, we did not consider relevant including lying HR too, since the information that we are trying to capture is already present in the model as represented by walking HR (and even better represented given the higher intensity of walking with respect to lying down). Moreover, if we were to include both HR parameters in the regression model, the sleeping HR parameter would be nonsignificant given the weaker link between sleeping HR and CRF with respect to walking HR and CRF, as shown by the lower correlation. The proposed approach could be used to provide more information about an individual's health without the need for laboratory infrastructure or specific tests. Building up on the proposed approach, new opportunities for applications targeted at inducing behavioral change could be developed, e.g., by creating a feedback loop between objectively measured physical activity and changes in CRF and associated reduced risk of disease.

ACKNOWLEDGMENTS

We thank Giuseppina Schiavone and Stefan Camps for support during data collection.

GRANTS

This work was funded by Holst Centre/imec.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

M.A., J.P., and O.A. conception and design of research; M.A. and P.C. analyzed data; M.A., P.C., G.P., and O.A. interpreted results of experiments; M.A. prepared figures; M.A. drafted manuscript; M.A., J.P., G.t.V., G.P., and O.A. edited and revised manuscript; G.t.V. performed experiments; G.P. and O.A. approved final version of manuscript.

REFERENCES

1. **Altini M, Penders J, Amft O.** Energy expenditure estimation using wearable sensors: a new methodology for activity-specific models. In: *Proceedings of the Conference on Wireless Health, WH '12*. New York, NY: ACM, vol. 1, p. 8.
2. **Altini M, Penders J, Amft O.** Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate. In: *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference*. Venice, Italy: IEEE, 2013, p. 65–72.
3. **Altini M, Penders J, Vullers R, Amft O.** Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. *IEEE J Biomed Health Inform* 99: 1, 2014.
4. **Altini M, Penders J, Vullers R, Amft O.** Personalized physical activity monitoring on the move. In: *Proceedings of the 4th Conference on Wireless Health, ser WH '13*. New York, NY: ACM, 2013, p. 8:1–8:2.
5. **Astrand PO, Ryhming I.** A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *J Appl Physiol* 7: 218–221, 1954.
6. **Bonomi AG, Plasqui G, Goris AH, Westerterp KR.** Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *J Appl Physiol* 107: 655–661, 2009.
7. **Brage S, Ekelund U, Brage N, Hennings MA, Froberg K, Franks PW, Wareham NJ, 2007.** Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol* 103: 682–692, 2007.
8. **Browning RC, Kram R.** Energetic cost and preferred speed of walking in obese vs. normal weight women. *Obesity Res* 13: 891–899, 2005.
9. **Cao ZB, Miyatake N, Higuchi J, Ishikawa-Takata K, Miyachi M, Tabata I.** Prediction of vo2max with daily step counts for Japanese adult women. *Eur J Pppl Physiol* 105: 289–296, 2009.
10. **Ebbeling CB, Ward A, Puleo EM, Widrick J, Rippe JM.** Development of a single-stage submaximal treadmill walking test. *Med Sci Sports Exerc* 23: 966–973, 1991.
11. **Esco MR.** Cross-validation of the polar fitness test TM via the polar f11 heart rate monitor in predicting $V_{O_2\,max}$. *J Exercise Physiol* 14: 31–37, 2011.
12. **Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE.** Prediction of functional aerobic capacity without exercise testing. *Med Sci Sports Exerc* 22: 863–870, 1990.
13. **Jetté M, Campbell J, Mongeon J, Routhier R.** The Canadian Home Fitness Test as a predictor of aerobic capacity. *Can Med Assoc J* 114: 680–682, 1976.
14. **Katzel LI, Bleecker ER, Colman EG, Rogus EM, Sorkin JD, Goldberg AP.** Effects of weight loss vs aerobic exercise training on risk factors for coronary disease in healthy, obese, middle-aged and older men: a randomized controlled trial. *J Am Med Assoc* 274: 1915–1921, 1995.
15. **Kuipers H, Verstappen F, Keizer H, Geurten P, Van Kranenburg G.** Variability of aerobic performance in the laboratory and its physiologic correlates. *Int J Sports Med* 6: 197–201, 1985.
16. **Lee DC, Artero EG, Sui E, Blair SN.** Mortality trends in the general population: the importance of cardiorespiratory fitness. *J Psychopharmacol* 24, *Suppl* 4: 27–35, 2010.
17. **Loimaala A, Huikuri H, Oja P, Pasanen M, Vuori I.** Controlled 5-mo aerobic training improves heart rate but not heart rate variability or baroreflex sensitivity. *J Appl Physiol* 89: 1825–1829, 2000.
18. **Mello RP, Murphy MM, Vogel JA.** Relationship between a two mile run for time and maximal oxygen uptake. *J Strength Cond Res* 2: 9–12, 1988.
19. **Minetti AE, Boldrini L, Brusamolin L, Zamparo P, McKee T.** A feedback-controlled treadmill (treadmill-on-demand) and the spontaneous speed of walking and running in humans. *J Appl Physiol* 95: 838–843, 2003.
20. **Nes BM, Janszky I, Vatten LJ, Nilsen T, Aspenes ST, Wisløff U.** Estimating vo2peak from a nonexercise prediction model: the hunt study, Norway. *Med Sci Sports Exerc* 43: 2024–2030, 2011.
21. **Noonan V, Dean E.** Submaximal exercise testing: clinical application and interpretation. *Phys Ther* 80: 782–807, 2000.
22. **Plasqui G, Westerterp KR.** Accelerometry and heart rate as a measure of physical fitness: cross-validation. *Med Sci Sports Exerc* 38: 1510–1514, 2006.
23. **Rennie KL, Hennings SJ, Mitchell J, Wareham NJ.** Estimating energy expenditure by heart-rate monitoring without individual calibration. *Med Sci Sports Exerc* 33: 939–945, 2001.
24. **Romero I, Grundlehner B, Penders J.** Robust beat detector for ambulatory cardiac monitoring. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. Minneapolis, MN: IEEE, 2009, p. 950–953.
25. **Rothney MP, Neumann M, Béziat A, Chen KY.** An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol* 103: 1419–1427, 2007.
26. **Santo AS, Golding LA.** Predicting maximum oxygen uptake from a modified 3-minute step test. *Res Q Exer Sport* 741: 110–115, 2003.
27. **Sartor F, Vernillo G, de Morree HM, Bonomi AG, La Torre A, Kubis HP, Veicsteinas A.** Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. *Sports Med* 43: 865–873, 2013.
28. **Tapia E.** *Using Machine Learning for Real-Time Activity Recognition and Estimation of Energy Expenditure* (PhD thesis). Cambridge, MA: Massachusetts Institute of Technology, 2008.
29. **Tonis T, Gorter K, Vollenbroek-Hutten M, Hermens H.** Comparing vo2max determined by using the relation between heart rate and accelerometry with submaximal estimated vo2max. *J Sports Med Phys Fitness* 52: 337–343, 2012.
30. **Uth N, Sørensen H, Overgaard K, Pedersen PK.** Estimation of VO2max from the ratio between HRmax and HRrest-the heart rate ratio method. *Eur J Appl Physiol* 91: 111–115, 2004.
31. **Vanhees L, Lefevre J, Philippaerts R, Martens M, Huygens W, Troosters T, Beunen G.** How to assess physical activity. How to assess physical fitness. *Eur J Cardiovasc Prevention Rehab* 12: 102–114, 2005.
32. **Weyand PG, Kelly M, Blackadar T, Darley JC, Oliver SR, Ohlenbusch NE, Hoyt RW.** Ambulatory estimates of maximal aerobic power from foot-ground contact times and heart rates in running humans. *J Appl Physiol* 91: 451–458, 2001.