Title:

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

Authors:

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

Authors' contribution:

- Marco Altini: study design, data analysis and paper writing
- Pierluigi Casale: support during data analysis
- Julien Penders: support during study design and paper writing
- Gabrielle ten Velde: support during data collection and paper writing
- Guy Plasqui: support during data collection and paper writing
- Oliver Amft: support during study design and paper writing

Affiliations:

[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

[4] University of Passau, Germany, and Eindhoven University of Technology, The Netherlands

Running head:

Estimating cardiorespiratory fitness in free-living

Address for correspondence:

Marco Altini, Eindhoven University of Technology, Den Dolech 2, 5612AZ, email: altini.marco@gmail.com, phone: 0031 6 46375742, fax: 0031 40 246 3120

Abstract:

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 use context-specific HR to estimate cardiorespiratory fitness (CRF). CRF of 51 participants was assessed by a maximal exertion test (VO_2max). Participants wore a combined accelerometer and HR monitor during a laboratory based simulation of activities of daily living and for two weeks 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 to 0.78 when including fat-free mass). Then, 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. RMSE was reduced from 354.7 ml/min 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.

New and noteworthy:

Many methods have been developed to estimate VO_2max using data collected under supervised laboratory conditions or following strict protocols. However, to the best of our knowledge, this is the first work, which proposes pattern recognition methods to contextualize heart rate (HR) in free-living and use context-specific HR to predict VO_2max. The proposed method does not require laboratory tests or specific protocols, showing error reductions up to 21% compared to VO_2max estimates derived using anthropometrics only.

Keywords:

Cardiorespiratory fitness, wearable sensors, heart rate, physical activity, context recognition

Introduction:

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 (10, 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 insights on an individual's health status, non-invasively, 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.VO_2 max) (30, 31). However, VO_2max measurements require medical supervision and can be risky for individuals where exercise until maximal exertion is contra-indicated. Despite the indubitable importance of CRF in health, measurements of VO_2max are therefore rare (21) and less risky submaximal tests have been developed. Non-exercise CRF estimation models use easily accessible measures such as age, gender and a self-reported physical activity level (14, 20). However, for individuals with similar anthropometric characteristics, CRF levels cannot be discriminated accurately. Alternatively, submaximal tests have been introduced to estimate VO_2max during specific protocols while monitoring HR at predefined workloads (5, 11). 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 VO_2max. However the need for laboratory equipment and the necessity to re-perform 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 sensors 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 VO_2max 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 VO_2max has been evaluated by different research groups (2, 23, 28, 30). 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 VO_2max 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 inter-individual differences in HR could be accounted for by surrogates of fitness such as measured or estimated sub-maximal 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 VO_2 at a certain exercise intensity cannot be exploited using only motion based sensors. Plasqui et al. (22) showed that a combination of average HR and physical activity over a period of 7 days correlates significantly with VO_2max. 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 sub-maximal 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 VO_2max 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 to similar relations between context-specific HR and VO_2max.

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.

Table 1: Participants' characteristics

Parameter	Mean ± SD
n	51 (24 male, 27 female)
Age (y)	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
VO_2max (ml/min)	3037.5 ± 671.6

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 read-out, 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 Li-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 tri-axial accelerometer (ADXL330) at 64 Hz. The ADXL330 accelerometer provides a ±3g 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 antero-posterior 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 more than 20%, as typically performed in clinical practice for heart rate 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 re-breathing except for the volume of the applied face-mask. Total flow was measured and converted to STPD values with a large dry-bellows flowmeter calibrated to 0.2% of used range by 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" rack units) and Servomex (Servomex 4100 and Servopro 5400 19" rack units) with a resolution ≤ 0.001% absolute. The analyzers are mounted separately to exclude both vibration and climate-variation as confounding factors. Ranges for the analyzers are set to 0-21% O2 and 0-1% CO2 yet only limited to 25% and 2.5% respectively. No specific smoothing is applied as the result is updated each 5 seconds while breathing is mechanically averaged in the ± 30 liter internal volume and the dilution gasflow, resulting in a time constant of 4.5 second at the 400 l/min setting. The calorimeter is validated by gas-infusion or burning fuel (methanol) over its full range (200 to 7000 ml.min-1) with a 1±2 % avg ± SD result. VO_2 max was reached when a plateau in VO_2 was observed and/or an RQ of 1.1 or higher. VO_2max was calculated as the highest average VO_2 over 30 seconds (6 consecutive values).

Study design:

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

- Laboratory protocols: participants reported at the lab on three separate days and after refraining from drinking, eating and smoking in the two hours before the experiment. Two laboratory protocols were performed, while the third 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, The Netherlands), to determine context-specific HR during activities of daily living simulated in laboratory settings. Activities included: lying down, sitting, sit and write, standing, cleaning a table, sweeping the floor, walking (treadmill flat at 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6 km/h) and running (treadmill flat at 7, 8, 9, 10 km/h). Activities were carried out for a period of at least 4 minutes.
    - The second protocol was a VO_2max test providing reference data for biking and CRF. VO_2max was determined during an incremental test on a cycle ergometer according to the protocol of Kuipers et al. (17). 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 bpm below the age-predicted maximal HR (208 – 0.7 x age) or the respiratory quotient exceeded 1, 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 hour. 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 VO_2max. 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 non-individualized and no laboratory data from the participant to be validated was used for model building. The procedure used for model building and evaluation is shown in Figure 1. For the beat detection we relied on methods developed by the research community in the past as these models are standard components that are already available in many sensor devices today. More details on the validation procedures are reported in the Statistics and performance measures Section. Context-specific HR in free-living was used in a multiple regression model to estimate VO_2max 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.

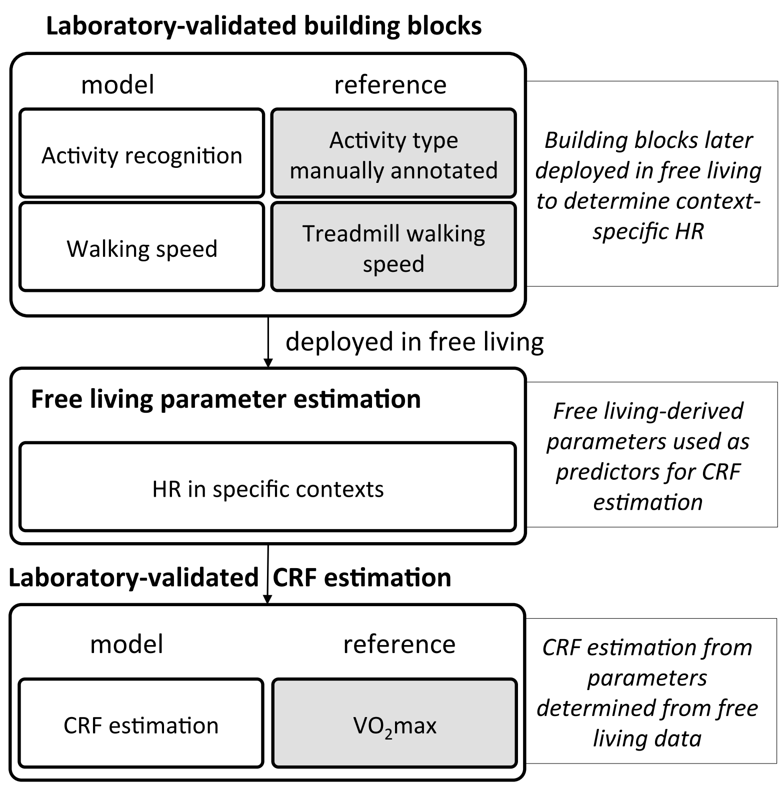

Figure 1: Block diagram of the proposed approach and validation procedure. Activity recognition and walking speed estimation models are built and validated using supervised laboratory recordings. Then, models are deployed in free-living. Activity recognition and walking speed estimation are used to determine HR in specific contexts in free-living. Finally, HR in specific contexts (e.g. HR while lying down or walking at a certain speed) are used as predictors for

VO2max estimation, effectively estimating 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 was used for model building, as described in the Statistics section. An example of activity recognition and walking speed estimation models output is shown in Figure 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. Secondly, the raw acceleration signal was processed to determine walking speed for activities recognized as walking. The acceleration signal was segmented in non-overlapping intervals of 5 seconds. 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, while 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 shown to be ideal in discriminating static gravitational acceleration and body motion, as shown in Fig. 4. Fig. 4 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, inter-quartile 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.

Table 2: Accelerometer features used for activity classification and walking speed estimation. N indicates the number of samples in a 5 seconds window, i.e. 160 (32 samples per second). LP and BP stand for low pass and band pass filtered data. Qn represents the nth quartile.

Feature name	Computation	Description

Mean of the absolute signal	$\frac{1}{N}\sum_{i=1}^{N}\lvert a_{BPi}\rvert$	Represents motion intensity independently of the axis or orientation, similarly to *activity counts*
Inter quartile 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	$\frac{1}{N}\sum_{i=1}^{N}(a_i - \mu)^2$ 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)

HR was extracted from RR intervals, and averaged over 15 seconds windows. Features for the multiple linear regression model used to estimate walking speed were: mean of the absolute signal, inter-quartile 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.

Table 3: Coefficients of the linear regression model used to estimate walking speed. During validation all models were evaluated using leave-one-participant-out cross-validation, the coefficients shown here include all data.

Parameter	Estimate coefficient	P value
Intercept	-1.000e+00	<2e-16
Mean of the absolute signal	9.936e00	<2e-16
Variance	- 2.963e00	<2e-16
Quartile (X axis)	3.256e00	<2e-16
Quartile (Y axis)	- 2.475e00	<2e-16
High frequency band signal power (X axis)	- 5.920e-04	<2e-16
High frequency band signal power (Z axis)	- 6.320e-04	<2e-16
Main frequency peak (X axis)	- 1.323e-01	<2e-16
Participant height	1.439e-02	<2e-16

Laboratory activities were grouped into six clusters to be used for activity classification. The six clusters were lying (lying down), sedentary (sitting, sit and write, 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 activities 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 activities 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. 2. We limited the features space to two dimensions to provide a visualization that is easily human readable. Fig. 2.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 low intensity activities (e.g. lying, sedentary), as show in Fig. 2.a. Fig. 2.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. 2c illustrate that the two shown features provide relevant information to discriminate the activities clusters. Hence already two features are sufficient to separate most - but not all - activity clusters in this study.

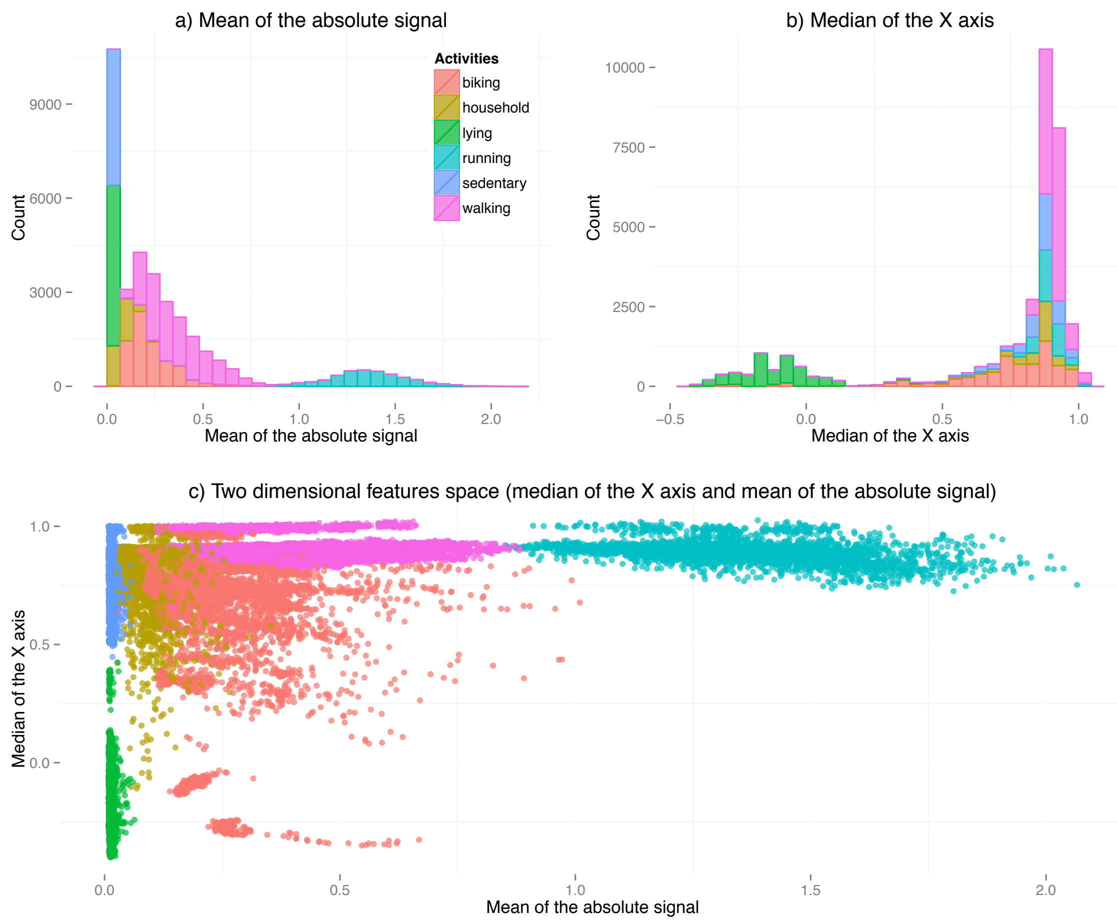

Figure 2. Example of extracted features and multi-dimensional features space used for activity classification of the six activity clusters included in this study. a-b) Histograms of two accelerometer features (mean of the absolute signal and median of the X axis). c) Two dimensional features space showing clear separations between most activity clusters.

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. 3. Fig 3.a. shows multiple example decision boundaries that separate the example data points, while Fig 3.b. shows the separating hyperplane that maximizes the margin to the example data points, as determined by the SVM.

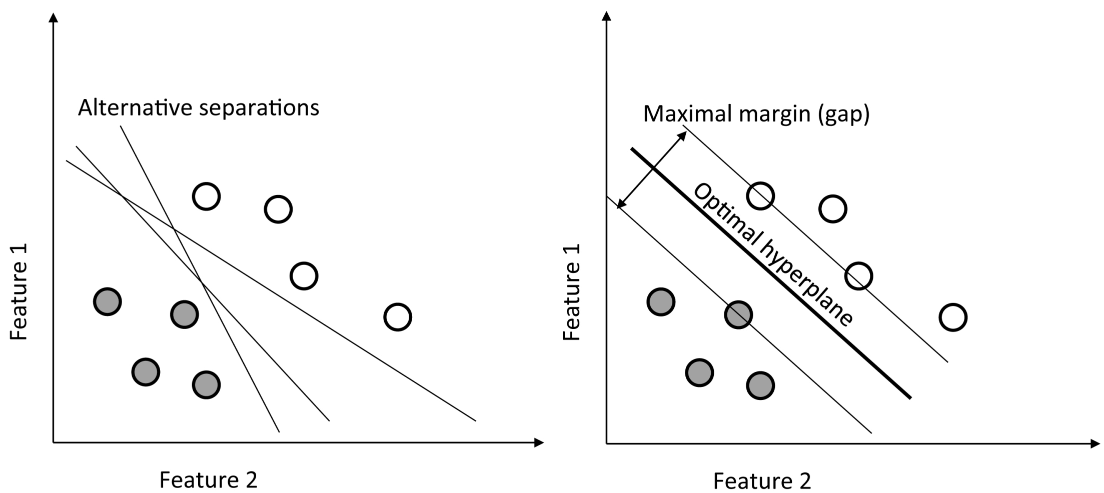

Figure 3. Example of linear decision boundaries to classify two classes. Example data points of the classes are illustrated in different gray tones. On the left; different example decision boundaries. On the right; optimal hyperplane obtained by maximizing the margin between the decision boundary and the closest data points of each class. Samples on the maximal margin lines are called support vectors.

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 VO_2max. We predicted VO_2max 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 was reported in previous studies between 5 and 6 km/h (5.3 km/h in (8) and 5 ± 0.8 km/h in (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 higher than 3 km/h and lower than 4 km/h as segments to be considered of an average walking speed of 3.5 km/h. Similarly, we selected data segments with detected speed higher than 5 km/h and lower than 6 km/h as segments to be considered of an average walking speed of 5.5 km/h.

Then, 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 VO_2max 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.

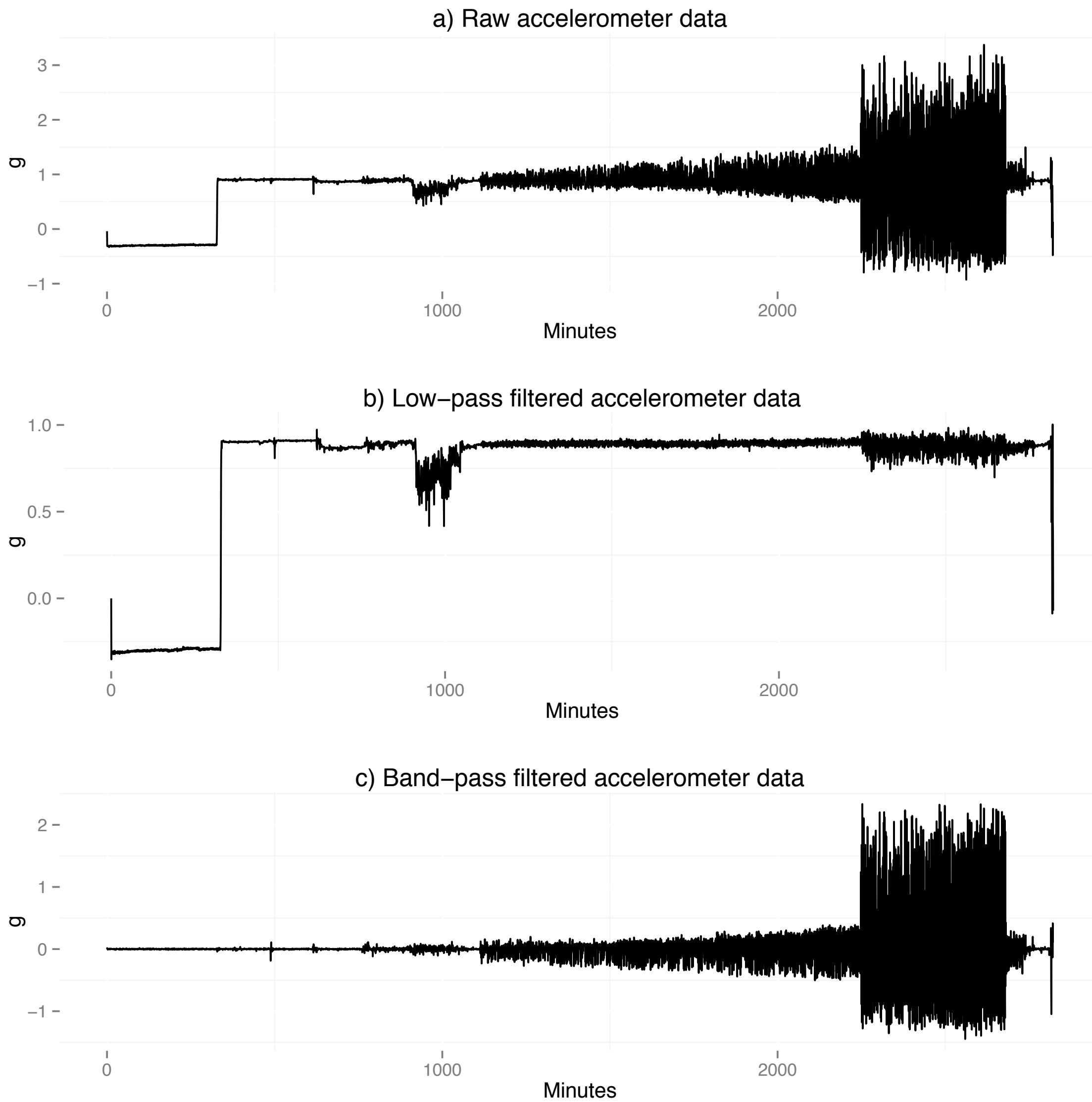

Figure 4: Raw accelerometer data (top), low-pass filtered data (center) and band-pass filtered data (bottom). The gravity component is isolated when using low-pass filtered data, as shown in the center plot. This information is particularly useful to distinguish postures. Band-pass filtered data isolates the accelerometer component due to body motion, showing increased values for higher intensity motions. Band-pass filtered data is particularly useful to distinguish ambulatory activities and walking speeds. Data were downsampled for visualization purposes.

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 was 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 km/h. The relation 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 standard deviation 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 as well as 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 VO_2max in ml/min 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 about 10 days per participant, including accelerometer and ECG data. Eighty-three hours of laboratory recordings including reference VO_2, VCO_2, acceleration, ECG and VO_2max 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. Fig. 5 shows an exemplary output of the walking speed and activity recognition models for one participant during 24 hours 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.

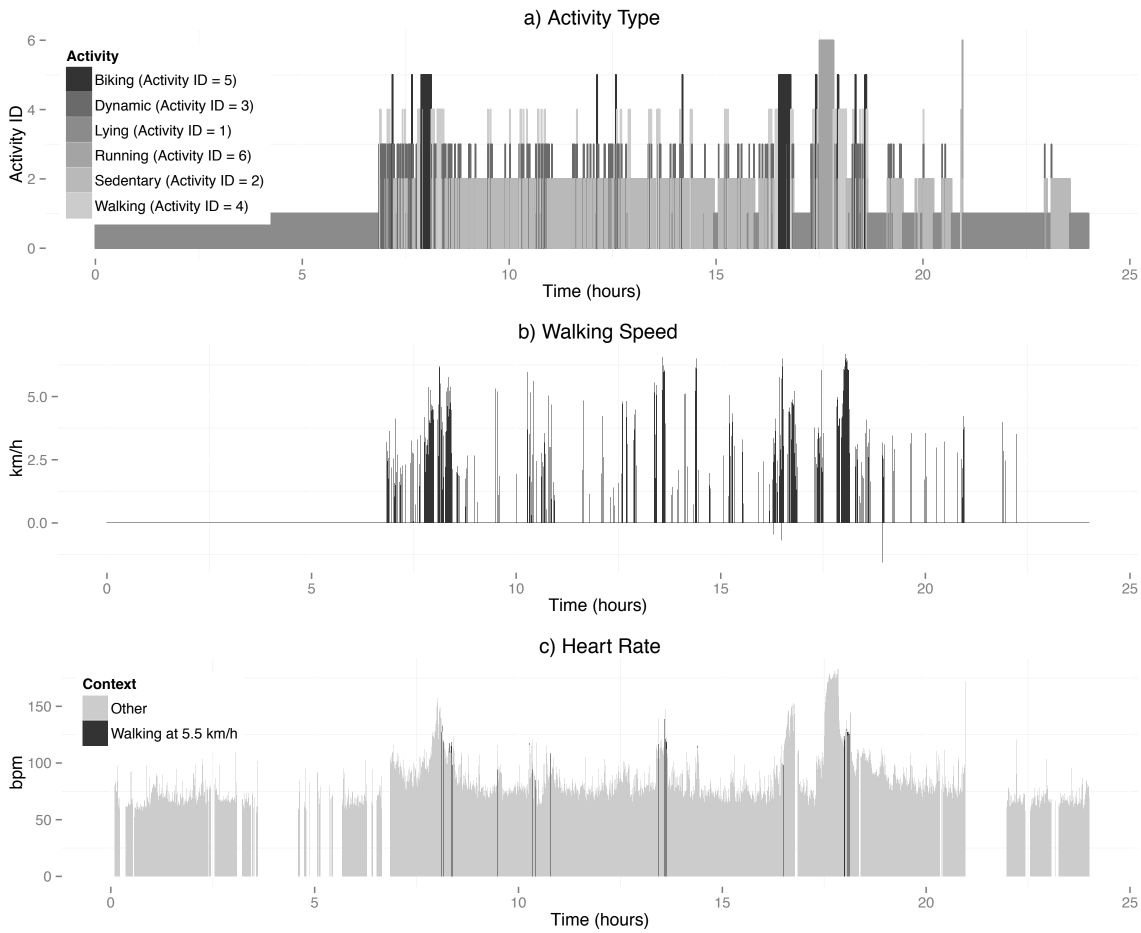

Figure 5: Exemplary output of the models used to contextualize HR in free-living in this work, for one participant. a) Recognized activity types. Commuting by bike, training (running), sleep and a mostly sedentary job during waking hours can be easily identified from this plot. b) Estimated walking speeds 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.

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 bpm for lying, 91.0 ± 15.3 bpm for walking at 3.5 km/h and 107.8 ± 17.7 bpm 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 as well as 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 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 the HR while walking at 3.5 km/h and p=0.0048 when including the HR while walking at 5.5 km/h).

Table 4: Multiple linear regression models for VO_2max estimation from activities of daily living simulated in laboratory settings. N = 49. For each predictor, detailed information (model coefficient, p-value) are indicated.

Model description	Predictors	R^2
Anthropometric characteristics only	Intercept (1431.686, p = 0.00322) Body weight (18.510, p = 0.00645), age (-1.888, p = 0.84595), sex (798.561, p = 7.48e-07)	0.64
Context-specific HR	Intercept (2849.294, p = 3.86e-05) HR while lying down in laboratory settings (-14.268, p = 0.00344), body weight (13.862, p = 0.02862), age (-8.170, p = 0.37335), sex (803.669, p = 1.20e-07)	0.69

	Intercept (3183.644, p = 5.67e-06), HR while walking at 3.5 km/h in laboratory settings (-13.636, p = 0.000496), body weight (14.921, p = 0.013229), age (-12.046, p = 0.183661), sex (777.430, p = 1.03e-07)	0.72
	Intercept (3367.865, p = 9.65e-07), HR while walking at 5.5 km/h in laboratory settings (-13.044, p=8.21e-05), body weight (15.234, p = 0.00856), age (-13.222, p = 0.13055), sex (754.772, p = 8.95e-08)	0.74

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, in percentage.

Classification results							
		Lying	Sedentary	Dynamic	Walking	Biking	Running
True Activities	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%

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 minutes per day walking, 11.9 minutes of which were at 3.5 km/h and 11.6 minutes 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 bpm in free-living and 66.2 bpm 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 bpm in free-living and 91.0 bpm 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 bpm in free-living and 107.8 bpm 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.

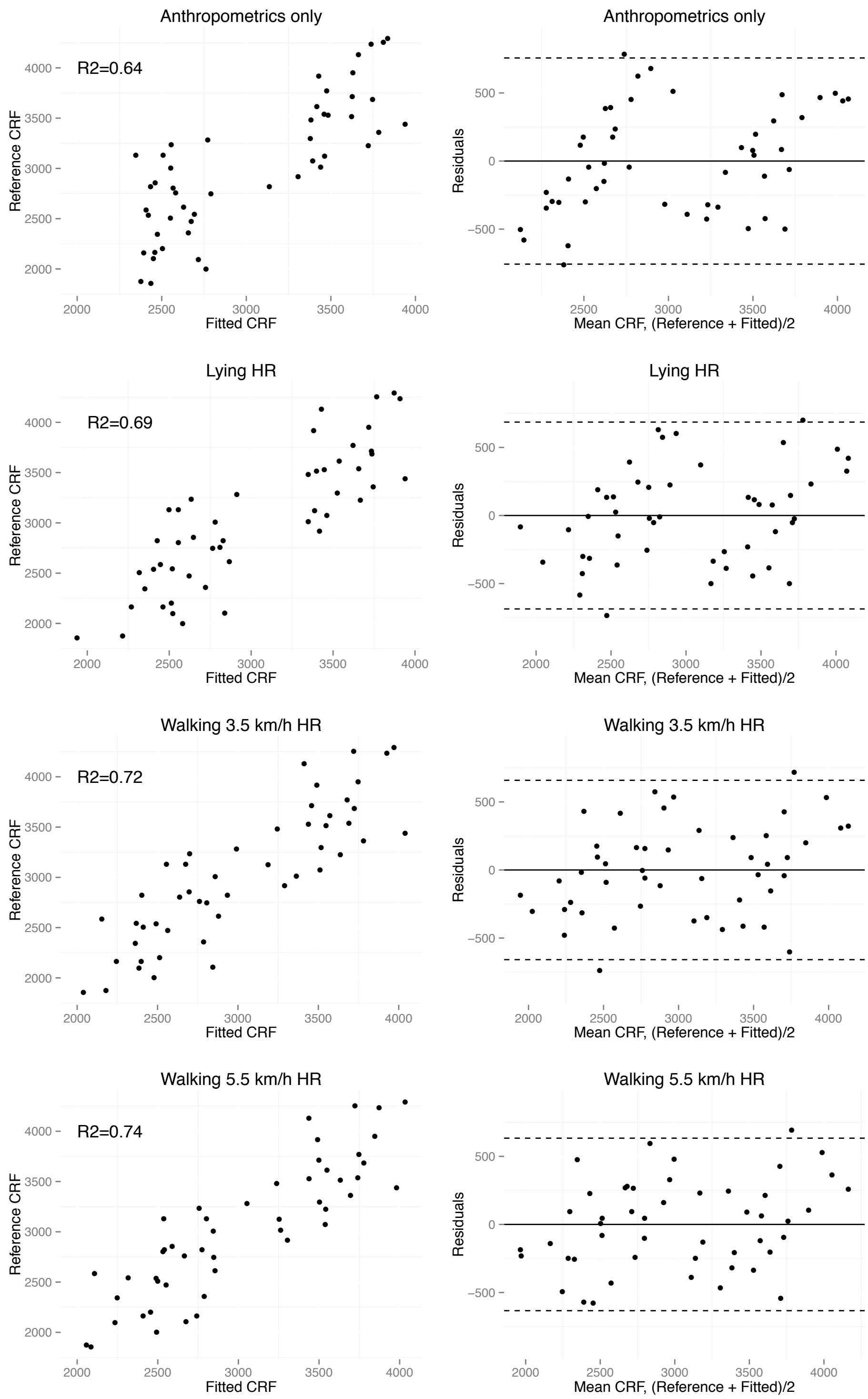

Figure 6. Accuracy of the prediction models for CRF estimation. Regression plots and Bland-Altman plots are shown for models using as predictors anthropometrics and context-specific HR during activities of daily living simulated in laboratory conditions. R^2 is also reported.

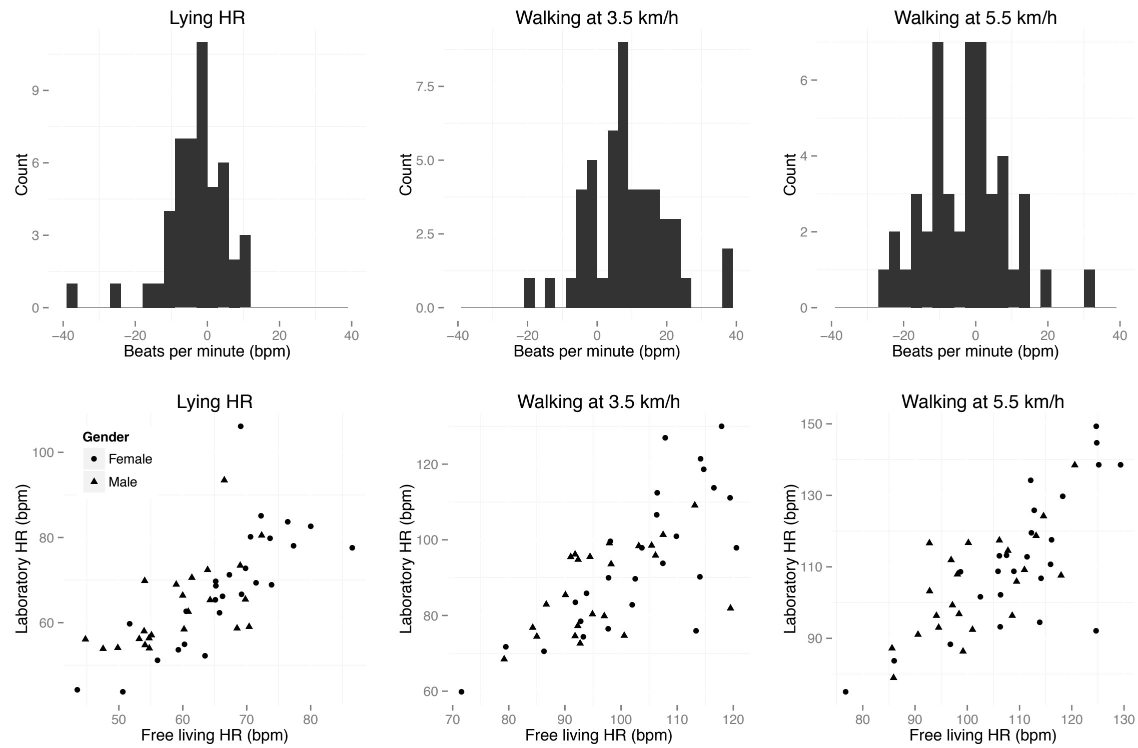

Figure 7. Top row: histograms of differences between context-specific HR in laboratory settings and free-living. Bottom row: scatterplots showing the relation between context-specific HR in laboratory settings and free-living.

CRF estimation from context-specific submaximal HR in free-living:

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

Table 6: Multiple linear regression models for VO_2max estimation from free-living data. N = 51. For each predictor, detailed information (model coefficient, p-value) are indicated.

Model description	Predictors	R^2
Anthropometric characteristics only	Intercept (1403.603, p = 0.00326), Body weight (19.531, p = 0.00355), age (-3.184, p = 0.73931), sex (803.869, p = 4.59e-07)	0.65
Context-specific HR	Intercept (2914.307, p = 6.31e-06), HR while lying down in free-living (-22.118, p = 0.000554), body weight (21.150, p = 0.000511), age (-9.027, p = 0.298184), sex (634.875, p = 1.32e-05)	0.73
	Intercept (4175.338, p = 2.19e-06), HR while walking at 3.5 km/h in free-living (-20.798, p = 0.000136), body weight (16.106, p = 0.005611), age (-20.240, p = 0.032176), sex (738.579, p = 1.55e-07)	0.74
	Intercept (4647.138, p = 1.03e-07), HR while walking at 5.5 km/h in free-living (-23.884, p = 7.03e-06), body weight (16.801, p = 0.0022), age (-21.322, p = 0.0156), sex (668.687, p=5.02e-07)	0.77

Cross-validation of VO_2max estimates:

VO_2max estimation models derived from free-living data were cross-validated using the leave-one-out technique. Results are reported in Table 7 and 8. Cross-validation of VO_2max 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 VO_2max 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.

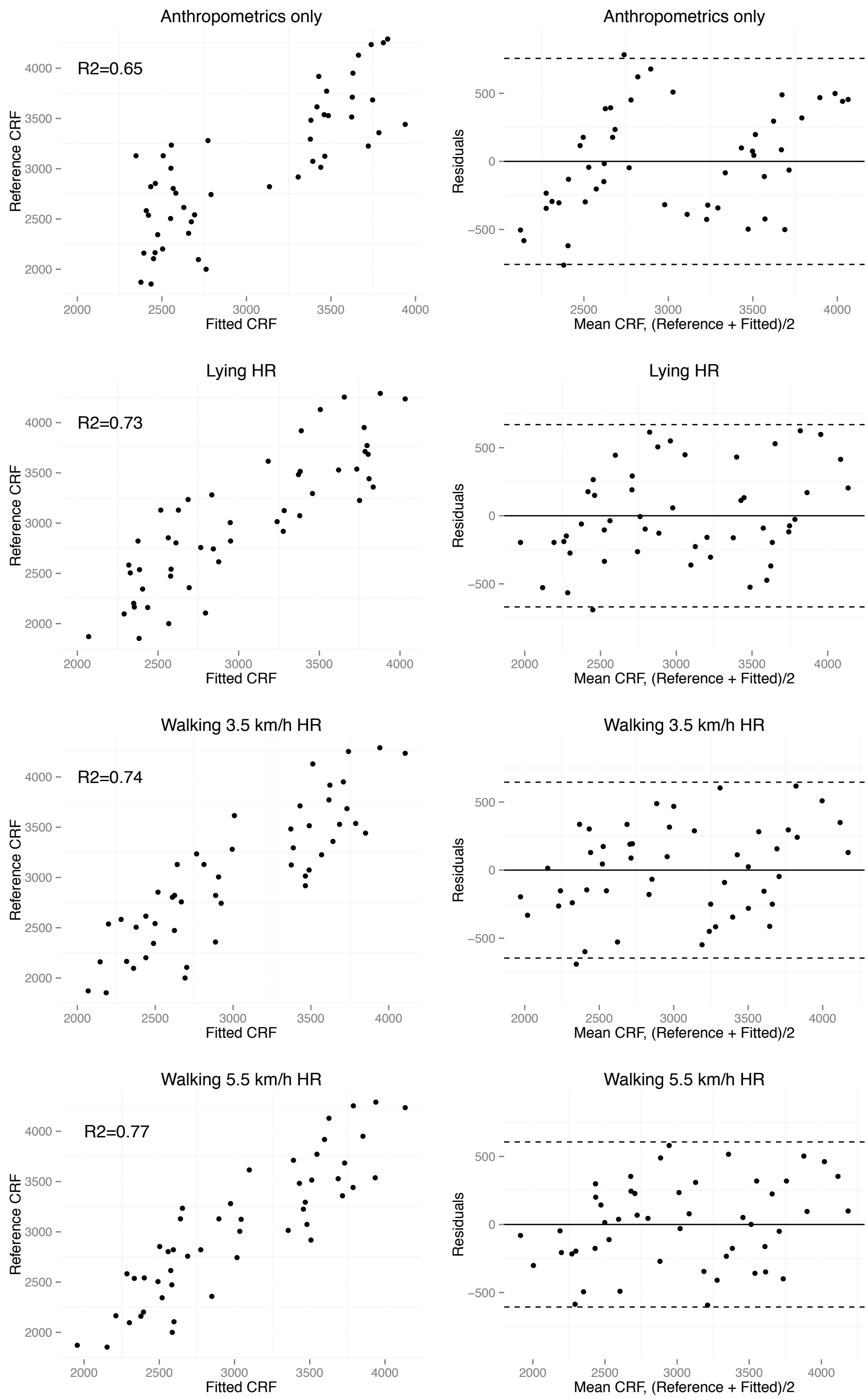

Figure 8. Accuracy of the prediction models for CRF estimation. Regression plots and Bland-Altman plots are shown for models using as predictors anthropometrics and context-specific HR in free-living. R^2 is also reported.

Table 7: Cross validation of multiple linear regression models for VO_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 weight, age, sex	358.3	0.66
Context-specific HR	HR while lying down in laboratory settings, body weight, age, sex	314.3	0.73
	HR while walking at 3.5 km/h in laboratory settings, body weight, age, sex	310.0	0.75
	HR while walking at 5.5 km/h in laboratory settings, body weight, age, sex	284.7	0.78

Table 8: Cross validation of multiple linear regression models for VO_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 weight, age, sex	354.7	0.67
Context-specific HR	HR while lying down in free-living, body weight, age, sex	309.4	0.75
	HR while walking at 3.5 km/h in free-living, body weight, age, sex	305.9	0.76
	HR while walking at 5.5 km/h in free-living, body weight, age, sex	281.0	0.79

Discussion:

In this work, we proposed a method to estimate VO_2max in free-living, without the need for laboratory tests or specific protocols. While many methods have been developed to estimate VO_2max 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 VO_2max during activities of daily living simulated in laboratory settings. Then we analyzed the correlation and relative differences between context-specific HR during activities simulated in the lab 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 VO_2max estimation using as predictors context-specific HR in free living provides accuracy comparable with laboratory derived models. In particular, RMSE for VO_2max estimation could be reduced up to 21% compared to 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 VO_2max, 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 VO_2max 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 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 VO_2max estimation during more intense activities, such as biking or running, compared to 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, in order to contextualized 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 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. Secondly, 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 VO_2max estimation (13, 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 minutes daily were spent at 3.5 km/h, while 11.6 minutes 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 minutes (e.g. the common 6-minutes walking test), we believe a total of 10 minutes daily is a sufficient amount of data for prediction of VO_2max, 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 confusions. 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 ml/min, 310.0 ml/min and 284.7 ml/min for lying, walking at 3.5 km/h and walking at 5.5 km/h were respectively) are similar to RMSE results obtained contextualizing HR using pattern recognition methods in free-living (309.4 ml/min, 305.9 ml/min 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 VO_2max similar to what we reported in laboratory settings. The inverse relation between HR at a certain workload and VO_2max is the key principle behind laboratory based submaximal CRF tests and this relation showed 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 VO_2max 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 VO_2max. Explained variance also increased, between 0.65 when anthropometrics characteristics only were used to estimate VO_2max, 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 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 bpm and 63.2 bpm 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 VO_2max estimation including fat free mass instead of body weight among the predictors resulted in higher accuracy, as expected and previously shown in 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, since the aim of our work is to provide VO_2max 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 VO_2max estimates: We also performed cross validation using subject independent models for VO_2max estimation as 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 is used in the models. Thus, the lower RMSE showed by VO_2max 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 VO_2max 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 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 literature on protocol-free VO_2max estimation. Previous studies aiming at estimating VO_2max 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 et al. (22), or requiring intense exercise such as running (32). Results for VO_2max 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's participants characteristics, for example body weight and VO_2max 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 VO_2max estimation. For some studies, e.g. 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 et al. on a cross-validation sample for his method showed that using as predictor HR divided by activity counts, a measure of motion intensity, VO_2max 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 predictor, we obtained results comparable to or better than previous free-living studies and are also comparable to what was reported using similar metrics in laboratory settings or while performing strict protocols (25). For example, ninety-two different VO_2max protocols were reviewed in a recent analysis by Sartor et al. (27). Additionally to the free-living studies here discussed, the authors suggested that many other sub-maximal tests could be performed in free-living, without laboratory infrastructure. However, most of these tests require intense activities and strict protocols, for example the most commonly used 2-mile run (Mello et al. (18), R^2 = 0.81), Canadian aerobic fitness test (Jette et al. (15), R^2 = 0.82), or YMCA (Santo et al. (26), R^2 = 0.56). 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 VO_2max could be continuously assessed longitudinally over time, and not only re-assessed 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 VO_2max (12). However, these studies typically reported low levels of accuracy (Esco et al. (12), R^2 = 0.29), showing that single measurements or "spot" measurements of physiological parameters and limited levels of context are insufficient for a reliable VO_2max estimate. A possible explanation for the better performance of the proposed approach compared to both single spot checks (12) 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 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 re-training 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 non-healthy 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 VO_2max to remain constant over a period of two weeks, since participants were not implementing changes to their lifestyle, and typical interventions to modify VO_2max are of much longer duration (e.g. 3 months to 1 year (16)). 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 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 VO_2max 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, as 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 VO_2max 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).

Conclusions: In conclusion, this work showed that contextualized HR in free-living can be used to provide VO_2max estimation with accuracy comparable to other methods relying on submaximal HR measured in laboratory settings. This is the first study utilizing pattern recognition methods to automatically contextualized 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 non-significant, 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. For example, by creating a feedback loop between objectively measured physical activity, and changes in CRF and associated reduced risk of disease.

Acknowledgments:

The authors would like to thank Giuseppina Schiavone and Stefan Camps for their support during data collection.

Disclosure:

This work was funded by Holst Centre/imec.

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, pages 1:8, New York, NY, USA, 2012. ACM.

[2] Altini M, Penders J, and Amft O. Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*, pp. 65-72. IEEE, 2013.

[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 Journal of Biomedical and Health Informatics, no. 99, p. 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, USA: ACM, 2013, pp. 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*, vol. 7, no. 2, pp. 218–221, 1954.

[6] Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Improving assess- ment 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, M.A., Froberg, K., Franks, P.W. and Wareham, N.J., 2007. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *Journal of Applied Physiology*, *103*(2), pp.682-692.

[8] Browning RC, Kram R. Energetic cost and preferred speed of walking in obese vs. normal weight women. *Obesity Research*, vol. 13,
no. 5, pp. 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. *European journal of applied physiology*, vol. 105, no. 2, pp. 289–296, 2009.

[10] Lee DC, Artero EG, Sui E, Blair SN. Review: Mortality trends in the general population: the importance of cardiorespiratory fitness. *Journal of Psychopharmacology*, vol. 24, no. 4 suppl, pp. 27–35, 2010.

[11] Ebbeling CB, Ward A, Puleo EM, Widrick J, Rippe JM. Development of a single-stage submaximal treadmill walking test. *Med Sci Sports Exerc*, vol. 23, no. 8, pp. 966–973, 1991.

[12] Esco MR, et al. Cross-validation of the polar fitness test TM via the polar f11 heart rate monitor in predicting vo2 max. *Age (yrs)* 24 (2011): 5-1.

[13] Loimaala, A., Huikuri, H., Oja, P., Pasanen, M., & Vuori, I. (2000). Controlled 5-mo aerobic training improves heart rate but not heart rate variability or baroreflex sensitivity. *Journal of Applied Physiology*, *89*(5), 1825-1829.

[14] Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. *Med Sci Sports Exerc*, vol. 22, no. 6, pp. 863–870, 1990.

[15] Jetté M, et al. The Canadian Home Fitness Test as a predictor of aerobic capacity. *Canadian Medical Association Journal* 114.8 (1976): 680.

[16] Katzel, L. I., Bleecker, E. R., Colman, E. G., Rogus, E. M., Sorkin, J. D., & Goldberg, A. P. (1995). 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. *Jama*, *274*(24), 1915-1921.

[17] Kuipers H, Verstappen F, Keizer H, Geurten P, Van Kranenburg G. Variability of aerobic performance in the laboratory and its physiologic correlates. *International journal of sports medicine*, vol. 6, no. 04, pp. 197–201, 1985.

[18] Mello RP, Murphy MM, Vogel JA. Relationship Between a Two Mile Run For Time and Maximal Oxygen Uptake. *The Journal of Strength & Conditioning Research* 2, no. 1 (1988): 9-12.

[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. *Journal of Applied Physiology*, vol. 95, no. 2, pp. 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*, vol. 43, no. 11, pp. 2024–30, 2011.

[21] Noonan V, Dean E. Submaximal exercise testing: clinical application and interpretation. *Physical Therapy*, vol. 80, no. 8, pp. 782–807, 2000.

[22] Plasqui G, Westerterp KR. Accelerometry and heart rate as a measure of physical fitness: cross-validation. *Med Sci Sports Exerc*, vol. 38, no. 8, pp. 1510–1514, 2006.

[23] Rennie, K.L., Hennings, S.J., Mitchell, J. and Wareham, N.J., 2001. Estimating energy expenditure by heart-rate monitoring without individual calibration. *Medicine and science in sports and exercise*, *33*(6), pp.939-945.

[24] Romero, I., Grundlehner, B., & Penders, J. (2009, September). Robust beat detector for ambulatory cardiac monitoring. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* (pp. 950-953). IEEE.

[25] Rothney, M. P., Neumann, M., Béziat, A., & Chen, K. Y. (2007). An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of applied physiology*, *103*(4), 1419-1427.

[26] Santo AS, Golding LA. Predicting maximum oxygen uptake from a modified 3-minute step test. *Research quarterly for exercise and sport* 74.1 (2003): 110-115.

[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 medicine*, vol. 43, no. 9, pp. 865–873, 2013.

[28] Tapia E. Using machine learning for real-time activity recognition and estimation of energy expenditure. In PhD thesis, MIT, 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. *The Journal of sports medicine and physical fitness*, vol. 52, no. 4, pp. 337–343, 2012.

[30] Uth, N., Sørensen, H., Overgaard, K., & Pedersen, P. K. (2004). Estimation of VO2max from the ratio between HRmax and HRrest–the heart rate ratio method. *European journal of applied physiology*, *91*(1), 111-115.

[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?. *European Journal of Cardiovascular Prevention & Rehabilitation*, vol. 12, no. 2, pp. 102–114, 2005.

[32] Weyand, P. G., Kelly, M., Blackadar, T., Darley, J. C., Oliver, S. R., Ohlenbusch, N. E., ... & Hoyt, R. W. (2001). Ambulatory estimates of maximal aerobic power from foot-ground contact times and heart rates in running humans. *Journal of Applied Physiology*, *91*(1), 451-458.