A 2-minute Fitness Test for Lifestyle Applications: The PhysioFit Task and Its Analysis based on Heart Rate

Neide Simões-Capela[1,2], Jan Cornelis[2], Giuseppina Schiavone[3] and Chris Van Hoof[1,2]
[1]*ESAT, KU Leuven, Kasteelpark Arenberg 10, Heverlee, Belgium*
[2]*IMEC, Kapeldreef 75, Heverlee, Belgium*
[3]*IMEC-NL, High Tech Campus 31, Eindhoven, The Netherlands*

Keywords: Health Related Fitness, Cardio-respiratory Fitness, Submaximal Fitness Test, Ruffier-Dickson Task, PhysioFit Task.

Abstract: Cardio-respiratory fitness (CRF) denotes the health of cardiorespiratory and musculoskeletal systems, thus being important to evaluate effects of (un)healthy lifestyles. Non-exhaustive submaximal fitness tests enable simple, fast, and inexpensive CRF assessment, in situations with low accuracy requirements. An example is the Ruffier-Dickson task (RD), consisting of 30 squats executed within 45 seconds, it estimates a CRF score from heart rate (HR) during the task. Squats, however, are not straightforward for subjects with poor fitness. To overcome this limitation, we developed the PhysioFit task (PF). It entails two minutes of stationary pedaling and employs HR for CRF estimation. PF outcomes were analyzed using RD as benchmark, according to HR changes during the task; CRF scores estimated with methods based on HR; correlation of CRF scores to body composition. The analysis relied on data from 28 subjects who executed both tasks. Although, HR variations during PF were lower relative to RD, PF produced significant changes in HR during pedaling and allowed for significant recovery after one minute. Significant agreement was found between tasks for two CRF scores, and both presented strong negative and positive correlations with fat and muscle percentage, respectively. Preliminary results show that PF is promising towards fast fitness assessments.

1 INTRODUCTION

Physical fitness describes how readily physical activities can be performed, something that can be defined in relation to health targets (i.e. health related) or to a specific athletic skill (i.e. skill related). The concept of health related fitness is the most pertinent for the general population as it quantifies diverse health aspects, namely body composition (BComp); cardiorespiratory endurance; muscular strength and endurance; and flexibility (McArdle, Katch, & Katch, 2015). Nonetheless, accurate fitness assessment is complex, expensive and has varied health contraindications, being mainly limited to athletes and specialized research. Simple and affordable alternatives exist that can fit a wide range of individuals, if a suboptimal accuracy is tolerated.

The work hereby presented is aimed at evaluating the potential of a 2-min pedaling task, the PhysioFit (PF), towards physical fitness assessment. The PF is compared to the Ruffier-Dickson (RD) task, a simple fitness test that relies on the execution of 30 squats to attain a fitness evaluation, based on heart rate (HR) during the exercise. The PF task offers an alternative for situations in which the former is not feasible (e.g. subjects with low weight or poor fitness). The analysis methods developed for the RD task are applied on PF task data, to test whether relevant fitness information can be extracted. The article is organized as follows. Chapter 1 introduces relevant concepts related to fitness and summarizes the state-of-the-art in data analysis. Chapter 2 details our methods for data collection and analysis. Chapter 3, 4 and 5 present results, discussion, and conclusion, respectively.

1.1 Body Composition

Body mass index (BMI) assesses the normalcy of a person's weight in relation to height, as in eq.1 (WHO, 2004).

$$\mathrm{BMI} = \mathrm{weight}/\mathrm{height}^2 \tag{1}$$

Analyzing BComp, a domain of health related fitness, provides a more comprehensive assessment,

Simões-Capela, N., Cornelis, J., Schiavone, G. and Van Hoof, C.
A 2-minute Fitness Test for Lifestyle Applications: The PhysioFit Task and Its Analysis based on Heart Rate.
DOI: 10.5220/0010234503770385
In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 5: HEALTHINF*, pages 377-385
ISBN: 978-989-758-490-9

quantifying relative amounts of fat, muscle, and bone in the body. The most convenient measurement method uses bioimpedance analysis. Normal ranges for each BComp component vary with age, gender, ethnicity and measuring device (McArdle et al., 2015).

1.2 Cardiorespiratory Fitness

Cardiorespiratory fitness (CRF) reflects the health of the cardiovascular, respiratory, and musculoskeletal systems. Thus, highly influencing the level to which everyday aerobic activities can be performed (Arena et al., 2007). Tough, its major interest lies on the inverse correlation to morbidity and mortality (Kodama et al., 2009). Several testing methodologies and descriptors are available for CRF assessment, and they are summarized next.

CRF testing comprises maximal and submaximal fitness tests. In maximal tests the exercise workload is incrementally increased until the test subject achieves volitional exhaustion. These tests provide the most accurate assessments. Though, they have limited applicability on the general population related to health contraindications (Thompson, Arena, Riebe, and Pescatello, 2013). Their widespread use is also limited by the requirements for specialized medical supervision, on-site emergency equipment, specific training, and elaborated protocols requiring expensive acquisition setups. Submaximal tests, on the other hand, do not require subjects to reach exhaustion. With less contraindications, they are suited for a wider range of individuals (e.g. children, elderly), and the varied acquisition protocols available meet different user requirements. Overall, submaximal tests present a convenient alternative to the maximal counterparts in situations with low CRF accuracy requirements (e.g. home, primary care, general research).

The golden standard for CRF assessment is the maximal oxygen uptake (VO_{2max}), achieved when consumed oxygen reaches a plateau, despite an increase in exercise load (i.e. when reaching exhaustion). It can be directly accessed by ventilatory gas analysis (Fletcher et al., 2013). CRF categorical classifications (e.g. very poor, poor, fair...) based on VO_{2max} are provided by the American College of Sports and Medicine (ACSM) (ACSM, 2014). Gender, age, height, body size/composition, training status and type of testing protocol all influence the VO_{2max} value (Fletcher et al., 2013). Height is especially important when the center of mass is displaced during the test protocol (McArdle, Katch, and Katch, 2015).

One CRF correlate that is easier to assess is HR. HR at rest (HR_{rest}) is a general indicator of wellness, while a decline in the HR response to submaximal exercise represents an enhancement in endurance. Also, the HR recovery pattern is a mortality predictor (ACSM, 2014). HR reaches its maximum (HR_{max}) approximately when VO_{2max} is achieved. For an increasing exercise load, HR increases linearly with oxygen consumption (VO_2). The relation holds during light to moderate workloads but may degrade for high workloads as VO_2 accelerates. The linearity of the HR-VO_2 relation has been used to predict VO_{2max} in submaximal tasks, by applying a linear regression to known points and extrapolating the relation up to a theoretical HR_{max}. Due to the assumptions put into this prediction, the estimated value is usually within 10-20% of the actual VO_{2max}. Some authors refer that this accuracy level is unacceptable for research, but can still be valuable in lifestyle applications (e.g. screening at the gym) (McArdle et al., 2015). Likewise, we argue that for non-fitness specific research, such estimates can be useful.

VO_{2max} has been derived from HR_{rest}, HR_{max} and weight as in eq.2 (N. Uth, 2005). The conversion to relative units (mL/min.kg) is required for comparison with guidelines and among subjects, which is achieved by multiplying by 1000/weight. In eq.2 the proportional factor (pf) takes different values for women and men: $14.5\text{x}10^{-3}$ l/min.kg and $15.3\text{x}10^{-3}$ l/min.kg, respectively. Theoretical HR_{max} ($HR_{max,th}$) can be estimated from eq.3 (Tanaka, Monahan, & Seals, 2001). Variations to this formulation are available, though no agreement exists on which is generally preferable (ACSM, 2014).

$$\mathrm{VO_{2max,Uth}} = {}^{\mathrm{weight} \times \mathrm{pf} \times \mathrm{HR_{max}}}/_{\mathrm{HR_{rest}}} \tag{2}$$

$$\mathrm{HR_{max,th}} = 208 - 0.7 \times \mathrm{age} \tag{3}$$

1.2.1 Ruffier-Dickson Task

The RD task is one of the simplest submaximal tasks found in the literature. It consists on resting for 5 min, performing 30 squats over a period of 45 s, and recovering for 5 min (Figure 2 (b)). The basic setup requires a stopwatch and a mat to lay down during rest and recovery. Bilateral squatting, involved in the task, primarily activates lower body musculature, but spinal and abdominal muscles are also engaged (Eliassen, Saeterbakken, & van den Tillaar, 2018). While early literature on the design of this task is not accessible online, recent studies compared the RD task results against maximal fitness tests and formulated predictive VO_{2max} models.

The RD task analysis traditionally relies on three discrete HR values: at rest (P0), right after exercise (P1) and 1 min into recovery (P2). A reference for the expected values is presented in Table 1. Such values are employed in calculating numerical fitness scores, such as the Ruffier index (Ri, eq.4) or the latter Ruffier-Dickson index (RDi, eq.5). The choice of factors in Ri and RDi is explained by De Mondernard et al. (De Mondenard, 1987), showing an ad-hoc process with weak empirical validation. Numerical outputs of Ri and RDi can be translated to fitness categories (e.g. excellent, good, fair...) as described in previous literature (De Mondenard, 1987)(Dah, 1991)(Sartor et al., 2016), though the classification ranges vary. The interest of Ri and RDi scores in fitness evaluation was re-evaluated in recent studies, that examined their relation to VO_{2max} measured during maximal tasks in healthy individuals.

Table 1: RD task: expected HR at rest, maximum steady state during exercise and 1 min into recovery (De Mondenard, 1987).

P0: Rest	P1: Adaption	P2: Recovery
P0<50 bpm: good Basal endurance. P0>80 bpm: poor basal endurance.	P1<2P0: good condition. P1>2P0: insufficient training.	P2≤P0: very good/good endurance. P2>P0+20: insufficient training.

$$\mathrm{Ri} = ((\mathrm{P0}+\mathrm{P1}+\mathrm{P2})-200)/10 \quad (4)$$

$$\mathrm{RDi} = ((\mathrm{P1}-70)+2(\mathrm{P2}-\mathrm{P0}))/10 \quad (5)$$

Sartor et al. (Sartor et al., 2016) shown that RDi alone should not be used to classify CRF levels in healthy subjects: the index shown low agreement (kappa=0.29) to ACSM CRF categorical levels and explained only 15% of the variability (adjusted r^2=0.15, sensitivity for good and fair=61%, specificity for poor=49%). Including RDi, age, gender (0=female, 1=male) and height on a multivariate model (eq.6), enhanced the agreement with ACSM levels (kappa=0.39) and the explained variability to 53% (adj. r2=0.53, sensitivity for good and fair=62%, specificity for poor=63%). These authors developed other models using HR values other than P0, P1 and P2. In our view, the increase in complexity did not justify the performance enhancement (adj. r^2=0.59, sensitivity for good and fair=64%, specificity for poor=62%, kappa=0.42), thus, those models are not detailed here. The models were developed on data from 81 healthy subjects (18 F, 63 M), with age ∈[18, 67] years old, height ∈[1.61, 1.88] m, weight ∈[52.5, 100] kg, and BMI ∈[18.8, 33.6] kg/m^2.

$$\begin{aligned}\mathrm{VO}_{2\mathrm{max,sartor}} = &-3.79 + 0.56\mathrm{gender} \\ &- 0.03\mathrm{age} + 4.53\mathrm{heigh} \\ &- 0.09\mathrm{RDi}\end{aligned} \quad (6)$$

Guo et al. (Guo et al., 2018) developed three models to predict VO_{2max}, respectively based on Ri, RDi and HR values (P0, P1, P2). Neither Ri (p = 0.06) nor RDi (p = 0.32) were significant predictors of VO_{2max}. The best model (eq.7) was found using P0, P1, P2, age, gender (0=female, 1=male,) and height (adj. r^2=0.64, sensitivity for good and fair=79%, specificity for poor=56%, kappa = 0.6). The models were developed on 40 healthy subjects (22 F, 18 M), age ∈[19, 60] years old, height ∈[1.57, 1.93] m, weight ∈[49.9, 121.6] kg, BMI ∈[18.6, 41.2] kg/m^2, P0 ∈[49, 98] beats per minute (bpm), P1 ∈[101, 184] bpm, and P2 ∈[56, 152] bpm.

$$\begin{aligned}\mathrm{VO}_{2\mathrm{max,guo}} = &\ 3.014 + 1.16\mathrm{gender} \\ &- 0.03\frac{\mathrm{P0}}{\mathrm{height}} \\ &+ 118.76\frac{\mathrm{P1}-\mathrm{P2}}{\mathrm{age}^3}\end{aligned} \quad (7)$$

To calculate their performance metrics, both Sartor et al. (Sartor et al., 2016) and Guo et al. (Guo et al., 2018) used three CRF classes (i.e. poor, fair and good), adapted from ACSM's classification for VO_{2max} during the Balke treadmill protocol (ACSM, 2014).

1.2.2 PhysioFit Task

The PF task is a submaximal task comprising 5 min of rest while sitting, 2 min pedaling and 5 min of sited recovery (Figure 1(a)). Pedaling is performed in upright seated position on a stationary bike with a fixed gear, with the objective of attaining and maintaining 35 km/h. The setup requires a chair, a stopwatch, and a stationary minibike (i.e. without upper limb support). The pedaling activity primarily activates lower body musculature and secondarily arm, abdominal and back muscles when upper limbs are used for support (So, Ng, and Ng, 2005). The muscle activation is a close match to the RD task, an important factor when comparing both tasks, as VO_{2max} values predicted from upper and lower body exercises have low correlation (McArdle et al., 2015).

The PF task was thought to cater for varied levels of fitness, accounting for some individuals being unable to perform or repeat complex movements (e.g. squats, step-up/down); while keeping the setup portable (i.e. excluding treadmills or a bicycle ergometers) and affordable (minibike prices range from 20-200 euros, depending on brand); and excluding tasks requiring the test subject to leave the controlled experimental environment to perform them (e.g. field walk, run tests). This task was first employed in psychophysiological research related to eating disorders, as a physical stressor (Simões-Capela, Schiavone, De Raedt, Vrieze, & Van Hoof, 2019). The aim was to weight the effect of physical activity on bio signals, when primarily studying the effects of mental stress on the body.

2 METHODS

2.1 Data

The PF and RD tasks were compared based on data from two studies, in which volunteers with varied levels of fitness completed both tasks.

Dataset 1 results from a pilot study designed to compare both tasks, following standard task protocols. The study was reviewed and approved by the medical ethics committee of *Ziekenhuis Oost-Limburg*. The study sample consisted of 13 subjects (7M, 6F) from a working population with age=30.4±7.2 years and BMI=22.3±4.3 kg/m^2 (mean ± s. dev.). All agreed to voluntarily participate and consented to the data collection after an explanation of the study procedures. All subjects were older than 18 years and working on a day desk job (i.e. excluding physical exertive jobs and shift works). The following constituted exclusion criteria: sensitive skin or known allergy to Ag/AgCl electrodes; inability to perform the protocol (e.g. limited mobility, respiratory illness, cardiovascular illness); acute illness (e.g. flu); pregnancy; and carrying implanted devices. The participants interfaced with three devices: 1) Health Patch (imec/Biotelemetry), a sensing node attached to an adhesive chest patch, used to continuously capture ECG; 2) HBF-516 (Omron), a full BComp monitor, to measure weight and estimate fat and muscle percentages based on bio-impedance analysis; 3) low-cost uncalibrated minibike (crivit, LIDL), used during the workout. The minibike was compared to a calibrated device (deskcycle, 3Dinnovations) to attest the accuracy of its displayed velocity, and an error of 10km/h was found (i.e. 35km/h displayed as 3.5km/h). This was taken in consideration during data collection. The tasks (Figure 1) were conducted in a dedicated study room under the supervision of trained researchers. The tasks were performed at the same time (between 3h and 5h pm) on consecutive days to avoid circadian changes. The task order was randomized. At the first contact the admission criteria were verified and background information (i.e. age, gender, height, weight, and BComp parameters) was collected. After applying the wearable sensors, the tasks took place as depicted in Figure 1, while a timed slideshow presentation with directions was shown on screen for reference. A screen recording was captured, to document any time diversions and account for them in the analysis.

Dataset 2 was originally dedicated to test the effect of diverse activities on bio-signal quality, and its methods include slight variations from Dataset 1. It was incorporated here to extend the study sample. The study was reviewed and approved by the medical ethics committee of *Universiteit Ziekenhuis Leuven*. The study sample consists of 15 subjects (5 M, 10 F) from a working population, with age=34.2±10.3 years and BMI=22.4±3.1 kg/m^2 (mean ± s. dev.). The admission criteria, the ECG acquisition device and bike setup were the same as in Dataset 1. The body analyzer was not employed. All procedures (Figure 2, with relevant tasks highlighted) were performed on a single 90 min study session between 8h and 12h am. In contrast to Dataset 1, there was no randomization of the task order. Weight and height were self-reported.

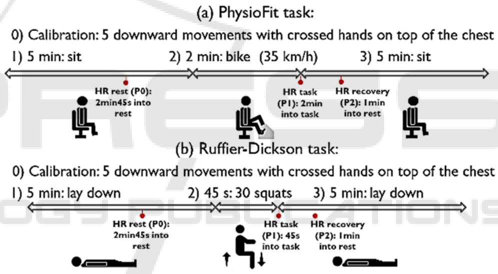


Figure 1: Task protocol for Dataset 1: (a) PF task and (b) RD task.

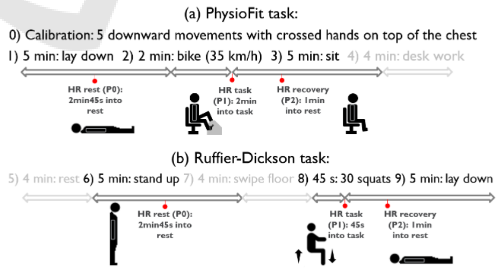


Figure 2: Task protocol for Dataset 2: (a) PF task and (b) RD task. Highlighted time slots were considered in the analysis.

2.2 Analysis

The analysis entailed: 1) pre-processing; 2) investigation of HR at P0, P1 and P2; 3) obtention of CRF scores using HR based models found in previous literature for the RD task; and 4) comparison of CRF scores to BComp. The previous steps were applied on data from both task, and results were compared within and between tasks. The analysis was carried out using MATLAB 2019b. All correlations were analyzed considering Cohen's guidelines (low correlation: $|\rho|\leq 0.3$; moderate correlation: $0.3<|\rho|<0.5$; strong correlation: $|\rho|\geq 0.5$).

ECG signals from each subject were truncated to the interval from start to end of each task. The start of each task was identified by the acceleration signature produced by the calibration procedure. Remaining phases were annotated based on the timings from the screen recordings. The R-peaks were identified in the ECG signal using an automatic beat detector (Romero, Grundlehner, & Penders, 2009). HR was calculated based on R-R intervals and converted to bpm. Each 1-min window was assessed for outliers (i.e. values outside the interval of mean HR ± 2.5 s. dev.) and these points were excluded.

Each of the relevant HR values was obtained by calculating the median of 15 seconds following the P0, P1 and P2 time points (cf. Figure 1 and Figure 2). Median HR values were used to reduce the effect of outliers. In both tasks it was investigated if changes in HR from rest to adaption, from adaption to recovery and from rest to recovery were significant. Among tasks, the HR values at each phase and the absolute HR variation from phase to phase were compared.

For both tasks, CRF scores were estimated based on 6 indices: Ri (eq.4), RDi (eq.5), $VO_{2max,sartor}$ (eq.6), $VO_{2max,guo}$ (eq.7), $VO_{2max,uth}$ (eq.2) and $VO_{2max,uth,th}$ (eq.2, eq.3). In $VO_{2max,uth}$, the HR_{max} was substituted by P1, in the expectation it would produce a proportional estimation. All outputs in units of L/min were translated to relative units of ml/kg.min. In both tasks the agreement between each pair of CRF indices was studied. Among tasks, the agreement among CRF values was tested.

The correlations of HR and CRF scores to BComp and BMI were investigated. Since only Dataset 1 includes information on BComp this analysis was limited to those 13 subjects.

3 RESULTS

This section includes results from the comparative analysis of HR and CRF scores within each task and among tasks. For most part of the analysis, Dataset 1 and 2 were treated as a single dataset, after visually verifying that both had a similar HR behavior (Figure 3). Background information of the study sample is summarized in Table 2.

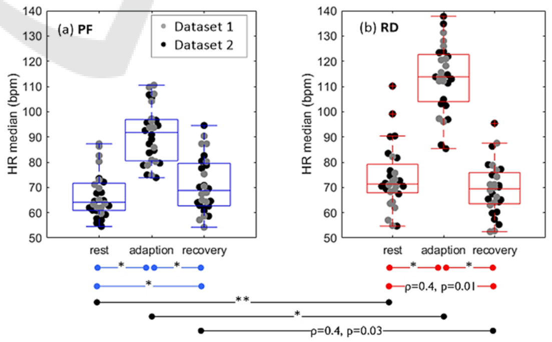
Table 2: Study sample: demographics and anthropometrics (mean ± s. dev. [max, min]).

	Male (N=12)	Female (N=16)	All (N=28)
Age, years	32.3 ± 5.6 [25, 44]	32.5 ± 11.2 [18, 52]	32.4 ± 9.1 [18, 52]
Height, m	1.83 ± 0.12 [1.69, 2.05]	1.65 ± 0.05[1.57,1.76]	1.73 ± 0.12 [1.57, 2.05]
Weight, kg	78.6 ± 19.0 [54.6,117.5]	59.7 ± 11.0 [43, 92]	67.8 ± 17.4 [43, 117.5]
BMI, kg/m²	23.2 ± 4.3 [18.4, 35.5]	21.7 ± 3.0 [16.4,29.7]	22.4 ± 3.6 [16.4, 35.5]

3.1 HR Intra and Inter-task

The distributions of P0, P1 and P2 are depicted in Figure 3, for both tasks. Based on the Kolmogorov-Smirnov (KS) normality test, all HR distributions are right-skewed, for such we used non-parametric statistics in the analysis. Data was not transformed to a normal distribution in order not to omit outliers. The Wilcoxon signed-rank test for dependent variables was employed to find significant differences. For non-significant differences, the presence of a linear relation was investigated based on Spearman's test.

In both tasks, there are significant differences between rest and adaption ($z=-4.6$, $p<<0.01$ for PF and $z=-4.6$, $p<<0.01$ for RD) and between adaption and recovery ($z=4.6$, $p<<0.01$ for PF and $z=4.6$ $p<<0.01$ for RD). Differences between rest and recovery are significant for the PF task ($z=-2.9$, $p<0.01$), but not for RD ($z=1.36$, $p=0.17$), in which case a significant moderate correlation is found ($\rho=0.4$, $p=0.01$). The absolute variations in HR_{median} from rest to adaptation are 27.6 bpm and 42.5 bpm, and from adaption to recovery are 23.0 bpm and 44.4 bpm, respectively for PF and RD task.



(c)	Median, PF	Ci, PF	Median, RD	Ci, RD
Rest, P0 (bpm)	64.1	[54.8,87.1]	71.3	[54.7, 107.9]
Adaption, P1 (bpm)	91.7	[74.0,110.3]	113.8	[85.6, 137.2]
Recovery, P2 (bpm)	68.7	[54.8,93.7]	69.4	[52.5, 93.8]

Figure 3: HR at rest (P0), adaption (P1) and recovery (P2): (a) PF task, (b) RD task, (c) statistics. Significant differences indicated with * ($p<0.01$) or ** ($0.01\leq$p-value<0.05), otherwise Spearman's ρ and p-value are shown. Ci: 95% confidence interval.

Across tasks, the HR is significantly different during rest (z=-2.4, p=0.02) and adaption (z=-4.6, p<<0.01), but not during recovery (z=1.0, p=0.3), presenting a significant moderate correlation in this case (ρ=0.4, p=0.03).

3.2 CRF Intra and Inter-task

All CRF scores' distributions are right skewed according to the KS normality test, hence non-parametric statistics were used in the analysis. The 6 CRF indices have different scales, and vary differently with an increasing level of fitness: Ri and RDi tend to -∞, while $VO_{2max,sartor}$, $VO_{2max,guo}$, $VO_{2max,uth}$ and $VO_{2max,uththeory}$ tend to +∞. In this case a test to compare medians is not appropriate. Thus, the relation among CRF indices' output was investigated using regression.

The linear regression between each pair of CRF indices (i.e. Ri-RDi, Ri-$VO_{2max,sartor}$, Ri-$VO_{2max,guo}$...) was calculated (Figure 4). For the PF task, four pairs of algorithms presented significant strong correlations: Ri-RDi (ρ=0.81, p>>0.01), Ri-$VO_{2max,uththeory}$ (ρ=-0.88, p>>0.01), RDi-$VO_{2max,uththeory}$ (ρ=-0.54, p>>0.01) and $VO_{2max,sartor}$-$VO_{2max,guo}$ (ρ=0.59, p>>0.01). For the same task, a moderate correlation was found for RDi-$VO_{2max,sartor}$ (ρ=-0.48, p=0.01). For the RD task, a significant strong correlation was found for five pairs: Ri-RDi (ρ=0.81, p>>0.01), Ri-$VO_{2max,uththeory}$ (ρ=-0.5, p=0.01), $VO_{2max,sartor}$-$VO_{2max,uththeory}$ (ρ=0.49, p=0.01), $VO_{2max,sartor}$-$VO_{2max,guo}$ (ρ=0.67, p>>0.01) and $VO_{2max,uth}$-$VO_{2max,uththeory}$ (ρ=-0.86, p>>0.01). For the same task,

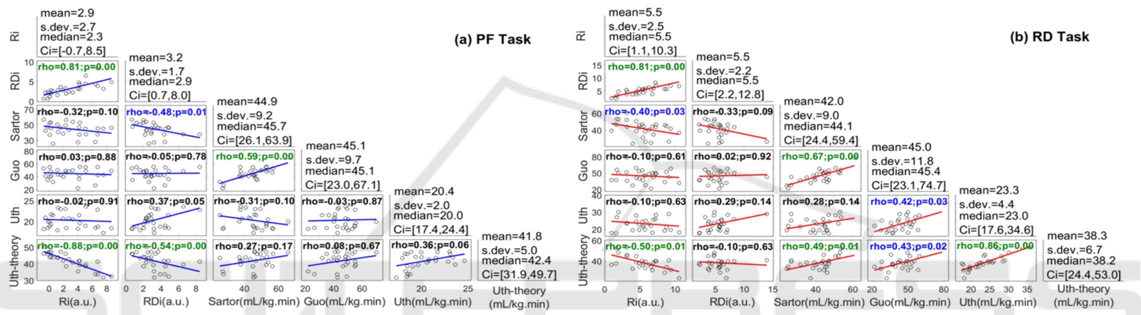


Figure 4: Linear relations between CRF indices for (a) PF task; (b) RD task. rho: Spearman's ρ, p: p-value, s. dev: standard deviation, Ci: 95% confidence interval. Significant correlations marked in green (p<0.01) or blue (0.01≤p-value<0.05).

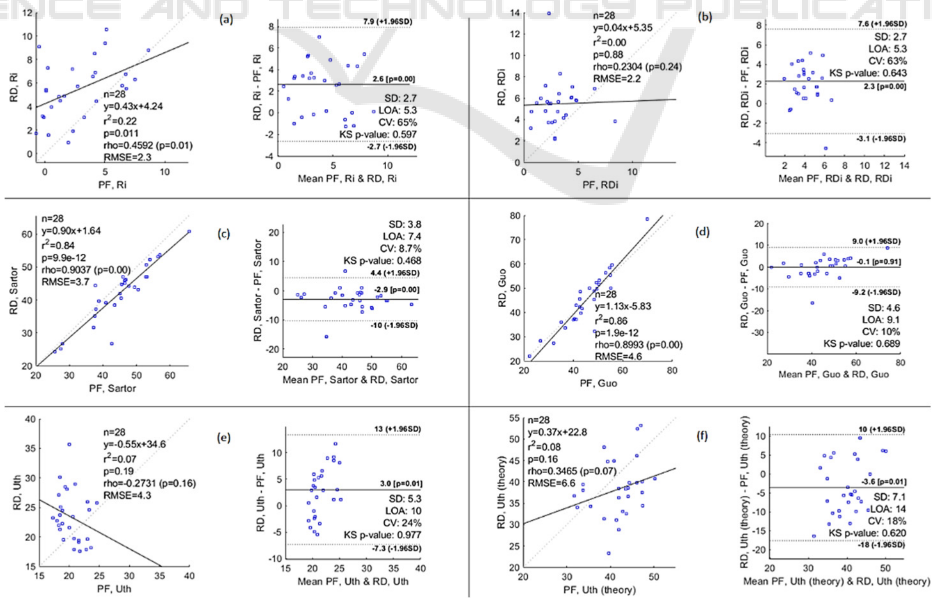


Figure 5: Linear regression of CRF scores from PF task against RD task and differences' plot for: (a)Ri, (b) RDi, (c) $VO_{2max,sartor}$, (d)$VO_{2max,guo}$, (e)$VO_{2max,uth}$ and (f)$VO_{2max,uth\ theory}$. n: number of points, y: regression formula, r^2: coef. of determination, p: Pearson's correlation p-value, rho (p): Spearman's ρ and p-value, RMSE: root mean squared error, SD: s. dev., LOA: limits of agreement (±1.96 s. dev.), presented as dashed lines, CV: coef. of variation (s. dev./mean in %), KS p-value: p-value for Kolmogorov-Smirnov test.

a moderate correlation was found for three pairs: Ri-$VO_{2max,sartor}$ (ρ=-0.40, p=0.33), $VO_{2max,guo}$-$VO_{2max,uththeory}$ (ρ=-0.43, p=0.02) and $VO_{2max,guo}$-$VO_{2max,uththeory}$ (ρ=0.42, p=0.03). Only three pairs of algorithms correlate well in both PF and RD tasks.

Regression analysis was used to investigate proportionality among CRF scores from different tasks. To understand if the relations found were significant the analysis of differences (Bland & Altman, 1999) was performed on the residuals (Figure 5), which is more robust to compare different acquisition methods than simple correlation. Three CRF indices shown significant agreement between tasks: Ri with moderate correlation (ρ=0.4, p=0.01); $VO_{2max,sartor}$ (ρ=0.9, p<<0.01) and $VO_{2max,guo}$ (ρ=0.9, p<<0.01) with strong correlation. All residuals arenormal (KS p-value>0.05), hence the results from regression are trustworthy. There is a systematic error between PF and RD for 5 CRF indices, as illustrated by the significant proportional bias on Ri (bias=2.6, p>>0.01), RDi (bias=2.3, p>>0.01), $VO_{2max,sartor}$ (bias=-2.9, p>>0.01), $VO_{2max,uth}$ (bias=3.0, p=0.01), and $VO_{2max,uththeory}$ (bias=-3.6, p=0.01). As for $VO_{2max,guo}$ (bias=-0.1, p=0.91) the bias is not significant. Ri and RDi have similar limits of agreement, both indicating a wide variability of the residuals. The $VO_{2max,sartor}$ (in comparison to $VO_{2max,guo}$, $VO_{2max,uth}$ and $VO_{2max,uththeory}$) presented the narrowest limits of agreement for the differences between tasks (LOA=7.4). The models with least dispersion of the residuals are $VO_{2max,sartor}$ (CV=8.7%) followed by $VO_{2max,guo}$ (CV=10%).

3.3 BMI and Body Composition

For Dataset 1 (N=13), correlations of $P_{0,1,2}$ and CRF scores to BMI, muscle and fat percentages were investigated based on Spearman's correlation test. Significant correlations are highlighted in Table 3.

Table 3: Correlation among CRF indices and anthropometrics for: (a) PF task, (b) RD task. Significant correlations marked with *(p<0.01) or **(0.01≤p-value<0.05).

	(a) PF task			(b) RD task		
	BMI	**Muscle %**	**Fat %**	**BMI**	**Muscle %**	**Fat %**
0	0.29	-.12	0.12	0.20	-.63**	0.48
P1	0.27	-.57**	0.46	-.35	-.02	-.15
P2	0.42	-.19	0.24	0.52	-.38	0.59
Ri	0.28	-.36	0.27	0.30	-.59**	0.54
RDi	0.31	-.48	0.43	0.07	-.18	0.20
$VO_{2max,sartor}$	-.56	0.92*	-.92*	-0.52	0.90*	-.90*
$VO_{2max,guo}$	-.66**	0.64**	-.84*	-.60**	0.66*	-.83*
$VO_{2max,uth}$	0.08	-.05	0.03	-.24	0.43	-.39
$VO_{2max,uththeory}$	-.22	0.26	-.19	-.16	0.70*	-.51

4 DISCUSSION

All HR distributions are right-skewed, which is a common find in HR literature, occurring whenever the sample as a subgroup of tachycardic subjects (Palatini, 1999).

For the intra-task comparison, we found that both tasks produce statistically and physiologically significant changes in HR from rest to adaption and both show significant recovery after adaption. In the RD task, HR at rest and HR during recovery have a significant moderate correlation. As for the PF task HR at rest and during recovery are significantly different.

For the HR inter-task comparison only HR during recovery agrees across tasks, presenting a significant moderate correlation. HR at rest can naturally change across measurements related to diet and activity prior to the measurement and acute changes in emotional state. Nonetheless, HR at rest is systematically higher for the RD task. In the current work, it is difficult to evaluate if body position (McArdle et al., 2015) is the reason for the difference, as Dataset 1 and Dataset 2 present varied resting positions. It is also possible that recovery is insufficient as the resting phase in Dataset 2 takes place after other activities. During adaption it is noted that pedaling leads to a significantly lower peak HR (less 22.1 bpm) than the squats, each producing a median variation in HR from rest to adaption of 27.6 bpm and 42.5 bpm, respectively. Such difference is unlikely due to body position, as this can only account for HR in PF (sitting) being systematically lower than RD (standing) by 1 bpm (McArdle et al., 2015) (Figure 3). During recovery there is a significant moderate correlation in HR across tasks, which could mean that the HR braking system acts to bring HR back to a baseline level, independently of the intensity of the physical stressor. Related to different body positions during recovery, HR in PF (sitting) should be systematically higher than RD (laying down). This should not invalidate the correlation found, though the ~1 bpm systematic error was not accounted in the calculation.

In the CRF intra-task comparison, we verified that not all models are consistent in the CRF score obtained for the same task (Figure 4). This does not appear to be a problem specific of the PF task: only 5 and 8 out 15 pairs of models agree for PF and RD, respectively. The discrepancies can be attributed to the different variables and their weight in each model.

For CRF inter-task comparison, two models shown to be consistent across tasks: $VO_{2max,guo}$ with no significant bias and $VO_{2max,sartor}$ systematically estimating higher fitness levels for the PF task (Figure

5). Unsurprisingly Ri and RDi shown poor agreement, as expected from their low rating as CRF predictors in previous literature (Sartor et al., 2016)(Guo et al., 2018). Finally, we verified that outputs from both $VO_{2max,guo}$ and $VO_{2max,sartor}$ agreed with other fitness indicator. With both presenting strong positive correlation with muscle percentage, and strong negative correlation with fat percentage (Table 3).

As the two different fitness tasks agree on the CRF scores obtained from two models that have been independently developed, and those scores agree with other fitness indicator (BComp), we illustrate the potential of our task for rough CRF estimation. Nonetheless, these are preliminary results and we are aware of the limitations of the current work. Dataset 2 presents design flaws related to the objectives this investigation, such as the incongruence of body positions with Dataset 1. We compare our task results to another submaximal task, while the correct approach towards validation is the comparison against a golden standard. Submaximal tests are especially useful for intra-subject comparison, over repeated measurements, which excludes reproducibility issues that are present across subjects. Our datasets present cross-sectional designs, preventing this analysis. Also, test-retest variability was not addressed. These limitations constitute points for further investigation.

5 CONCLUSIONS

We propose the PhysioFit, a simple 2-min pedaling task for fitness assessment, suited for subjects with low fitness level. We show that it induces a significant change in HR. We identify two models from previous literature (Sartor et al., 2016) (Guo et al., 2018) that can be used to analyze it, and obtain fitness scores based on HR during the task. CRF scores obtained from both models shown strong agreement with body composition indices. We reckon that this task is no match for settings requiring high accuracy assessments. Though, it has potential for rough fitness indexation in lifestyle and wellbeing applications (e.g. routine health checkups, tracking training progress or diet) or in non-fitness specific research studying human physiology (e.g. psychophysiology). With this work we intend to inspire the periodical monitoring of fitness levels in individuals who only casually engage in physical activity, be it in research studies, in the general practitioner's office, at home or in the work environment.

ACKNOWLEDGEMENTS

The authors acknowledge their gratitude to Emma Laporte for a preliminary literature review on fitness tasks; Erika Lutin and Christophe Smeets for reviewing the study materials; Luc Hons and Pieter Vandervoort for clinical supervision; and Leen Tordeurs for data management.

REFERENCES

ACSM. (2014). *ACSM's Guidelines for Exercise Testing and Prescription 9th Ed. 2014*.

Arena, R., Myers et al. (2007). Assessment of Functional Capacity in Clinical and Research Settings. *Circulation*, *116*(3), 329–343.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160.

Dah, C. (1991). Evaluation de l'aptitude physique. Intérêt, méthodes et application pratique. *Médecine d'Afrique Noire*, *38*(10), 681–687.

De Mondenard, J. J. (1987). Test des flexions de Ruffier-Dickson. *Ann. Kinésithér.*, (14), 381–388.

Eliassen, W., Saeterbakken, A. H., & van den Tillaar, R. (2018). Comparison of Bilateral and Unilateral Squat Exercises on Barbell Kinematics and Muscle Activation. *International Journal of Sports Physical Therapy*, *13*(5), 871–881.

Fletcher, G. F. et al. (2013). Exercise standards for testing and training. *Circulation*, *128*(8), 873–934.

Guo, Y. et al. (2018). A 3-minute test of cardiorespiratory fitness for use in primary care clinics. *PLoS ONE*, *13*(7), 1–11.

Kodama, S. et al. (2009). Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *Journal of the American Medical Association*, *301*(19), 2024–2035.

McArdle, W. ., Katch, F. I., & Katch, V. L. (2015). *Exercise Physiology: Energy, Nutrition and Human Performance, 8th edition*.

Palatini, P. (1999). Need for a Revision of the Normal Limits of Resting Heart Rate. *Hypertension*, *33*(2), 622–625.

Romero, I., Grundlehner, B., & Penders, J. (2009). Robust beat detector for ambulatory cardiac monitoring. *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 950–953.

Sartor, F. et al. (2016). A 45-second self-test for cardiorespiratory fitness: Heart rate-based estimation in healthy individuals. *PLoS ONE*, *11*(12).

Simões-Capela, N., Schiavone, G., De Raedt, W., Vrieze, E., & Van Hoof, C. (2019). Toward Quantifying the Psychopathology of Eating Disorders From the

Autonomic Nervous System Perspective: A Methodological Approach. *Frontiers in Neuroscience*, *13*(JUL), 1–12.

So, R. C. H., Ng, J. K.-F., & Ng, G. Y. F. (2005). Muscle recruitment pattern in cycling: a review. *Physical Therapy in Sport*, *6*(2), 89–96.

Tanaka, H., Monahan, K. D., & Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, *37*(1), 153–156.

Thompson, P. D., Arena, R., Riebe, D., & Pescatello, L. S. (2013). ACSM's New Preparticipation Health Screening Recommendations. *Current Sports Medicine Reports*, *12*(4), 215–217.

Uth, N. (2005). Gender difference in the proportionality factor between the mass specific V̇O2max and the ratio between HRmax and HR rest. *International Journal of Sports Medicine*, *26*(9), 763–767.

WHO. (2004, June 26). Body mass index. Retrieved June 26, 2020, from https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi