

Article

The Tyme Wear Smart Shirt is Reliable and Valid at Detecting Personalized Ventilatory Thresholds in Recreationally Active Individuals

Aaron H. Gouw¹, Gary P. Van Guilder¹, Gillian G. Cullen, Lance C. Dalleck^{1*}

¹ Western Colorado University, Recreation, Exercise & Sport Science, Gunnison, Colorado, USA;
aaron.gouw@western.edu; gvanguilder@western.edu; gillian.cullen@western.edu; ldalleck@western.edu

* Correspondence: ldalleck@western.edu; Tel.: +01-970-943-7132

Abstract: The aim of this study was to determine the extent to which the Tyme Wear smart shirt is as reliable and valid in detecting personalized ventilatory thresholds when compared to the Parvo Medics TrueOne 2400. In this validation study, 19 subjects were recruited to conduct two graded exercise test (GXT) trials. Each GXT trial was separated by seven to ten days of rest. During the GXT, gas exchange and heart rate data were collected by the TrueOne 2400 (TRUE) in addition to the ventilation data collected by the Tyme Wear smart shirt (S-PRED). Gas exchange data from TRUE was used to detect VT1 and VT2. TRUE and S-PRED VT1 and VT2 were compared to determine the reliability and validity of the smart shirt. Of the 19 subjects, data from 15 subjects were used during analysis. S-PRED exhibited excellent ($ICC > 0.90$) reliability for detection of VT1 and VT2 utilizing time point and workload and moderate ($0.90 > ICC > 0.75$) reliability utilizing heart rate. TRUE exhibited excellent reliability for detection of VT1 and VT2 utilizing time point, workload, and heart rate. When compared to TRUE, S-PRED appears to underestimate the VT1 workload ($p > 0.05$) across both trials and heart rate ($p < 0.05$) for trial 1. However, S-PRED appears to underestimate VT2 workload ($p < 0.05$) and heart rate ($p < 0.05$) across both trials. The result from this study suggests that the Tyme Wear smart shirt is less valid but is comparable in reliability when compared to the gold standard. Moreover, despite the underestimation of S-PRED VT1 and VT2, the S-PRED detected personalized ventilatory thresholds will provide an adequate training workload for most individuals. In conclusion, the Tyme Wear smart shirt provides easily accessible testing to establish threshold-guided training zones but does not devalue the long-standing laboratory equivalent.

Keywords: aerobic threshold, anaerobic threshold, exercise prescription, threshold detection, threshold-based training

1. Introduction

The development and use of consumer wearable technology has grown a considerable amount in the past five years. In 2016, the global market for wearables was valued at \$6.18 billion USD. In 2019, the market was valued at \$32.63 billion USD [1]. The global market size for wearable technology is expected to grow by 15.9% from 2020 to 2027 [1]. Typically, the goal of wearable technology has been to easily provide previously unknown biometric information regarding the user's physiology or lifestyle behavior. This information can be used to track fitness, guide training, or provide general health data such as heart rate, heart rate variability, oxygen saturation, and many other metrics. While the majority of wearable technology is typically used by consumers, there are wearable technologies that are used in the medical field with clinical populations. Even as the adoption of wearable technology continues to grow, there are continuing challenges associated with cost, poor data representation, lack of usability, and excessive information [2]. However, the greatest challenges that appears to plague wearable technologies are the reliability and validity of the biometric data and thus, the accuracy of the analysis and conclusions reported by these wearable devices.

In the context of wearable technology, reliability is defined as the consistency of the data collected from a measurement, which is also known as test-retest reliability [3]. To measure and test the reliability of the wearable, multiple measurements must be taken and compared to evaluate the ability for the wearable to consistently produce measurements. Validity is defined as the extent to which the data collected from a measurement accurately represents the reality of the variable(s) intended to be measured. In order to test validity, the wearable technology must be compared to the gold standard. For example, a smart watch that measures heart rate would have the validity of the heart rate measurement compared to an ambulatory electrocardiogram in order to determine the validity of the wearable. In addition, to support the analysis of reliability and validity of wearable technology, previous studies have utilized key statistical tests. The statistical analyses include: (1) Bland-Altman analysis and plots; (2) interclass correlation coefficient; (3) typical error [4].

While there are many physiological measures that can be measured to determine and support training related to exercise performance, ventilatory thresholds have been thoroughly utilized in both performance and clinical populations and supported with extensive evidence of their effectiveness to guide training and measure performance [5-7]. Determination of ventilatory thresholds is a function of minute ventilation (VE) and volume of oxygen consumed (VO_2) or VE and volume of carbon dioxide produced (VCO_2). The ventilatory thresholds, ventilatory threshold 1 and 2, will vary depending on the individual and correspond to specific exercise intensities or workload and demarcate shifts in exercise metabolism. Using these measurements, there are three primary ways that the ventilatory thresholds can be calculated: (1) ventilation curve method; (2) V-slope method; (3) ventilatory equivalent method [8].

With the continued advancement in hardware and software development, devices can now measure and collect physiological data that were once only possible in the laboratory environment. An example of this unique measurement in wearable technology is VE. The collection of VE data allows for the ability to determine ventilatory threshold in its users [9]. Tyme Wear has developed a smart shirt and app that purportedly measures VE to determine ventilatory thresholds. The data can then be used to provide training recommendations based upon the first and second ventilatory thresholds as well as performance prediction for running races such as the marathon. As with all wearable technologies, Tyme Wear should be tested for reliability and validity through comparison to the gold standard in order to be considered a viable product for the consumer.

The purpose of this study was to determine the reliability and validity of the Tyme Wear smart shirt compared to the Parvo Medics TrueOne 2400 in determining ventilatory thresholds in recreationally active men and women. As such, the question to be addressed: To what extent is the Tyme Wear smart shirt reliable and valid in determining ventilatory thresholds when compared to the gold standard?

2. Materials and Methods

Potential subjects were recruited via email or word of mouth. Subjects were Western Colorado University (WCU) students and staff as well as members of the Gunnison, Colorado community. Inclusionary criteria included recreationally active persons, male or female, between the ages of 18 and 65 years of age. Exclusion criteria included pulmonary and cardiovascular conditions as well as any physical conditions that may be agitated or exacerbated due to a running, maximal graded exercise test (GXT). Informed consent was provided to all subjects and acknowledged by all subjects prior to participation in the study. This study was reviewed and approved by the Western Colorado University Institutional Review Board [HRC2017-02-03-R77].

2.1 Experimental design

The experimental flow chart outlining the study design is shown in Figure 1. This validation study was conducted on the WCU campus in the High Altitude Performance Lab (HAP Lab) in Gunnison, Colorado, United States (2348 meters). The study consisted of two trials of a GXT that were separated by seven to ten days to allow for adequate rest between sessions. Exact rest days were determined based on scheduling availability of the researcher and subject. Two days prior to the initial GXT in trial 1, subjects were provided pre-test instruction and allowed to keep a copy of the instructions as a reminder for the second test. Pre-test questionnaire were completed prior to all tests to screen subjects for readiness. Subjects whose GXT did not meet the criteria of completion were given the option to: 1) complete an additional GXT 7-10 days post current test; 2) choose to no longer participate in the study.

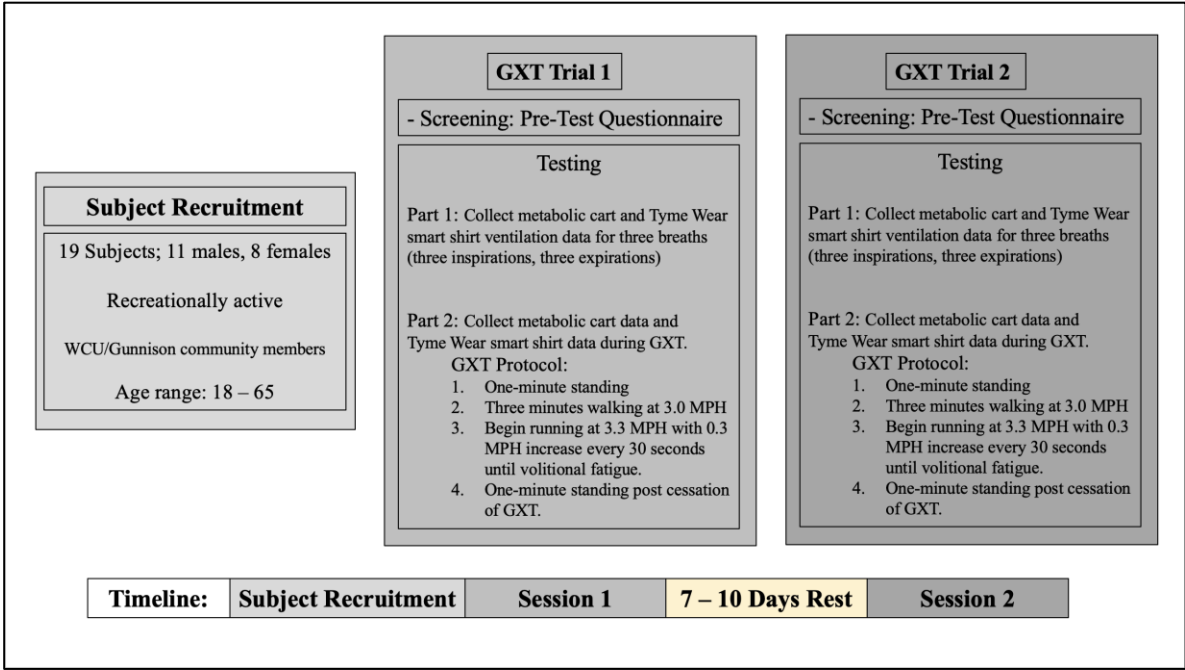


Figure 1. Experimental flow chart outlining study design.

2.2 Procedures

Pre-test preparation

Subjects were given a pretest questionnaire that determined the subject’s GXT readiness. The questionnaire was developed based upon variables that could influence the outcome of the GXT. Furthermore, this questionnaire allowed the subjects to replicate preparatory conditions (exercise, diet, etc.) for the second GXT session. If subject met all requirements for testing, they were prepared for the GXT. In order to collect heart rate data, subjects were fitted with a chest strap heart rate monitor (WearLink, Polar USA, Worcester, MA). Subjects were then provided a smart shirt (Tyme Wear smart shirt, Tyme Wear, Boston, MA) that was fitted according to the sizing guides provided by Tyme Wear. The Tyme Wear pod was subsequently attached to the hub of the smart shirt (refer to Figure 2 for external layout of shirt).



Figure 2. External Layout of Tyme Wear Smart Shirt and Pod.

Anthropometric Data Collection

Prior to the GXT, subject's height (cm) and weight (kg) were measured using a digital stadiometer (WB-3000, Tanita Corporation of America, Arlington, IL). Age was also recorded during this time.

Graded Exercise Test (GXT)

The protocol for the GXT was developed by Tyme Wear. Directions for the GXT was administered via a smart phone app and used to guide increases in exercise intensity during the GXT. Once the subject was connected to the metabolic cart (TrueOne 2400, Parvo Medics, Salt Lake City, UT) and connection between the app and smart shirt was confirmed, the GXT was initiated. All GXTs occurred on a motorized treadmill (Fitnexus Fitness Equipment, Dallas, TX). Heart rate was continuously measured and collected via the TrueOne2400 with a Polar heart rate sensor antenna (GymLink, Polar USA, Worcester, MA). Additional heart rate data was measured and collected separately on two polar heart rate monitor watches (Polar FS1, Polar USA, Worcester, MA) at 30 second intervals as redundancy in the event the TrueOne 2400 failed to collect heart rate data or to substitute anomalies in the heart rate data collected by the TrueOne2400. Prior to the GXT, breathing data from three full breaths (three full expirations and three full inspirations) was collected by the smart shirt and TrueOne2400. The GXT protocol began with one minute of standing rest while ventilatory data was collected by the smart shirt and gas exchange data collected by the TrueOne 2400. Subsequently, the subject was required to walk, as a warmup, at 3.0 mph for three minutes. Following this warm-up, the subject was required to begin running at 3.3 mph and running speed was increased 0.3 mph every 30 seconds until volitional fatigue. Then, the subject was asked to stand upright for one additional minute after cessation of the GXT while the smart shirt and metabolic cart continued to collect ventilation and gas exchange data, respectively.

Criteria for VO_2max Test

The criteria of for a true VO_2max test was established for this study according to previous research [10]. The following criteria were used to determine whether a subject reached VO_2max : 1) a respiratory exchange ratio (RER) of 1.10 or greater for men or women aged 20-49 and 1.05 or greater for men or women aged 50-64; 2) plateau of oxygen consumption less than or equal to $150 \text{ mL O}_2 \cdot \text{min}^{-1}$; 3) a rating of perceived exertion greater than eight (on a scale of 1 – 10; converted from Borg scale).

Identification of Ventilatory Thresholds Using Gas Exchange Data

The gold standard VT1 and VT2 were identified from gas exchange data through use of visual inspection (Wasserman et al., 1964). Visual assessment was completed by two trained exercise physiologists. In the situation there was disagreement, assessments were reevaluated until agreement was reached. This method for assessment and determination is consistent with past work [11-12]. Data averaged at 15-second intervals were plotted according to the appropriate variables for VT1 (VE/VO₂) and VT2 (VE/VCO₂). VT1 was identified based on the criteria of an increase in VE/VO₂ with VE increasing non-linearly in addition to the nonconcurrent increase in VE/VCO₂. VT2 was identified based on the concurrent increase of VE/VO₂ and VE/VCO₂.

2.3 Statistical analyses

Parameters (timepoint detection, heart rate, and workload) at VT1 and VT2 for the smart shirt (S-PRED) and TrueOne 2400 (TRUE) were compared using paired t-test to test for significant differences. A p-value less than .05 rejects the null hypothesis and suggests a statistically significant difference between groups. A p-value greater than .05 accepts the null hypothesis suggests a nonsignificant difference between groups. Analyses were conducted using the Microsoft Excel software (Microsoft Excel, Microsoft Corporation, Redmond, WA).

To determine the reliability of the Tyme Wear smart shirt, predicted VT1 and VT2 values from the two trials of each subject were compared to the equivalent data derived from the TrueOne 2400. To establish the reliability of the Tyme Wear smart shirt, the following statistical analyses were used: coefficient of variation (CV), typical error (TE), and intraclass correlation coefficient (ICC). Ranges for ICC are interpreted as excellent (ICC>0.9), moderate (0.90>ICC>0.75), fair (0.75>ICC>0.50), and poor (0.5>ICC) [13]. Reliability analysis was conducted using the Microsoft Excel software (Microsoft Excel, Microsoft Corporation, Redmond, WA).

To determine the validity of VT1 and VT2 determined by the Tyme Wear smart shirt, VT1 and VT2 determined by gas exchange data from the TrueOne 2400 were compared to the predicted VT1 and VT2 values determined by the ventilation data from the smart shirt. To compare the validity of the Tyme Wear smart shirt to the TrueOne2400, Bland Altman 95% limits-of-agreement were used to quantify the agreement (bias \pm random error (1.96 \times SD)) (Bland and Altman, 1986). Analyses and plots were conducted and created with the GraphPad Prism software (Prism, GraphPad Software, San Diego, CA).

3. Results

3.1 Subjects

Nineteen recreational active men and women volunteered for this study. Of the nineteen subjects tested, four subjects' data were excluded due to equipment error or data loss during testing. All subjects met the criteria for a valid VO₂max test. Relevant characteristics are described in Table 1

Table 1. Demographics and anthropometrics of subjects.

Subjects	Age (years)	Height (cm)	Body Mass (kg)	VO ₂ max (mL·kg ⁻¹ ·min ⁻¹)	Peak Heart Rate (bpm)
Total (n=15)	27.7 \pm 8.4	170.4 \pm 8.4	67.8 \pm 9.5	48.4 \pm 10.5	184.6 \pm 8.9
Male (n = 8)	29.0 \pm 9.5	173.1 \pm 8.5	73.1 \pm 8.1	52.8 \pm 12.4	183.8 \pm 11.1
Female (n = 7)	26.1 \pm 7.5	167.3 \pm 7.8	61.6 \pm 7.1	43.4 \pm 5.1	185.5 \pm 5.6

^a Values are mean \pm SD.

3.2 Reliability

Measures of reliability are described in Table 2 and Table 3 for TRUE and S-PRED, respectively. Reliability measures include CV, TE, ICC. Furthermore, the measures for reliability at VT1 and VT2 are presented across the three parameters: time point, heart rate, and workload. Table 2 shows that TRUE has excellent reliability in detecting VT1 and VT2 utilizing time point, heart rate, and workload. Similarly, Table 3 shows that S-PRED has excellent reliability in detecting VT1 and VT2 utilizing time point and workload. However, Table 3 shows that S-PRED has moderate reliability in detecting VT1 and VT2 utilizing heart rate.

Table 2. Measures of reliability for TRUE parameters: CV, TE, and ICC.

Parameters	CV (%)	TE	ICC
TRUE VT1			
Time	8.0	0.25	0.95
HR	3.2	0.35	0.90
Workload	6.8	0.25	0.95
TRUE VT2			
Time	5.3	0.22	0.96
HR	2.4	0.29	0.93
Workload	4.9	0.22	0.96

Units: Time (minutes), HR (bpm), Workload (mph).

Table 3. Measures of reliability for S-PRED parameters: CV, TE, and ICC.

Parameters	CV (%)	TE	ICC
S-PRED VT1			
Time	9.1	0.34	0.91
HR	3.9	0.58	0.78
Workload	6.9	0.33	0.91
S-PRED VT2			
Time	4.6	0.17	0.98
HR	2.8	0.48	0.84
Workload	4.3	0.18	0.97

Units: Time (minutes), HR (bpm), Workload (mph).

3.3 Validity

Averages for time point, workload, and heart rate values at VT1 and VT2 during GXT trials 1 and 2 for TRUE and S-PRED are presented in Table 4. Paired t-tests revealed significant mean differences for heart rate at VT1 ($p = 0.02$) and VT2 ($p = 0.02$) between TRUE and S-PRED during trial 1. Additionally, there were significant mean differences between TRUE and S-PRED for workload ($p = 0.02$) and time point detection ($p = 0.02$) at VT2 during trial 1. During trial 2, there were significant mean differences between TRUE and S-PRED time point detection ($p = 0.01$), workload ($p = 0.01$), and heart rate ($p = 0.02$) for VT2. Other TRUE versus S-PRED parameter comparisons were not significant ($p > 0.05$).

Table 4. Time point, workload, and heart rate at VT1 and VT2 for TRUE versus S-PRED during trial 1 and trial 2.

Parameter	Trial 1		Trial 2	
	TRUE	S-PRED	TRUE	S-PRED
VT1				
Time point (min)	6.3 ± 1.9 ^a	6.0 ± 1.6	6.5 ± 2.2	6.1 ± 1.6
Workload (mph)	4.6 ± 1.2	4.4 ± 0.9	4.6 ± 1.3	4.4 ± 1.0
Heart rate (bpm)	130.8 ± 10.9	126.8 ± 9.7 *	129.7 ± 13.3	125.4 ± 9.6
VT2				
Time point (min)	10.5 ± 3.0	9.1 ± 2.6 *	10.5 ± 3.0	9.3 ± 2.5 *
Workload (mph)	7.1 ± 1.8	6.6 ± 1.6 *	7.1 ± 1.8	6.4 ± 1.5 *
Heart rate (bpm)	163.5 ± 13.6	158.2 ± 8.9 *	162.3 ± 14.1	154.0 ± 10.89 *

^a Values are mean ± SD; * denotes significant between-group difference.

Furthermore, Table 5 presents the maximum, average, and minimum heart rate as a percent of achieved peak heart rate at VT1 and VT2 estimated by TRUE and S-PRED for the cohort. Paired t-test suggests significant mean differences between TRUE and S-PRED in the average % of max heart rate for VT1 in trial 1 ($p = 0.02$) as well as VT2 in trial 1 ($p = 0.01$) and trial 2 ($p = 0.02$). There was no significant difference between TRUE and S-PRED in the average % of max heart rate at VT1 for trial 2 ($p > 0.05$).

Table 5. VT1 and VT2 as a percent of peak heart rate.

Parameter	Trial 1		Trial 2	
	TRUE	S-PRED	TRUE	S-PRED
VT1				
Average % of Max	71 ± 5 ^a	69 ± 5*	70 ± 7	68 ± 5
Maximum HR %	81	78	85	77
Minimum HR %	63	58	57	57
VT2				
Average % of Max	88 ± 5	86 ± 3*	88 ± 7	83 ± 5*
Maximum HR %	96	90	99	88
Minimum HR %	78	81	72	76

^a Values are mean ± SD; * denotes significant between-group difference.

Similarly, Table 6 presents the maximum, average, and minimum workload as a percent of maximum workload at VT1 and VT2 estimated from TRUE and S-PRED for the cohort. In Table 6, paired t-test suggests significant mean differences between TRUE and S-PRED in the average % of max workload speed for VT2 in trial 1 ($p = 0.01$) and trial 2 ($p = 0.01$). There were no significant differences between TRUE and S-PRED in the average % of max workload at VT1 for trial 1 and trial 2 ($p > 0.05$).

Table 6. VT1 and VT2 as a percent of achieved maximum workload speed (mph).

Parameter	Trial 1		Trial 2	
	TRUE	S-PRED	TRUE	S-PRED
VT1				
Average % of Max	48 ± 5 ^a	46 ± 2	49 ± 8	46 ± 4
Maximum mph %	60	50	64	53
Minimum mph %	40	40	38	38
VT2				
Average % of Max	74 ± 7	69 ± 4*	75 ± 9	66 ± 5*
Maximum mph %	88	79	89	76
Minimum mph %	63	60	56	59

^a Values are mean ± SD; * denotes significant between-group difference.

3.4 Bland-Altman plots and analyses

Bland-Altman plots indicating the mean differences in GXT VT1 and VT2 time point detection between TRUE and S-PRED and levels of agreement with 95% CIs are illustrated in Figure 3 and Figure 4. For trial 1 VT1 time point detection, the mean difference between TRUE and S-PRED was 0.3 ± 0.9 minute (95% CI, -1.4 to 2.0 min). For trial 2 VT1 time point detection, the mean differences between TRUE and S-PRED was 0.4 ± 1.4 minute (95% CI, -2.2 to 3.1 min). For trial 1 VT2 time point detection, the mean differences between TRUE and S-PRED was 0.82 ± 1.46 minute (95% CI, -2.0 to 3.7 min). For trial 2 VT2 time point detection, the mean differences between TRUE and S-PRED was 1.2 ± 1.8 minute (95% CI, -2.4 to 4.8 min).

Bland-Altman plots indicating the mean differences in GXT VT1 and VT2 workload between TRUE and S-PRED and levels of agreement with 95% CIs are illustrated in Figure 5 and Figure 6. For trial 1 VT1 workload, the mean difference between TRUE and S-PRED was 0.2 ± 0.5 mph (95% CI, -0.8 to 1.2 mph). For trial 2 VT1 workload, the mean difference between TRUE and S-PRED was 0.3 ± 0.8 mph (95% CI, -1.3 to 1.8 mph). For trial 1 VT2 workload, the mean difference between TRUE and S-PRED was 0.5 ± 0.9 bpm (95% CI, -1.2 to 2.2 mph). For trial 2 VT2 workload, the mean differences between TRUE and S-PRED was 0.8 ± 1.1 mph (95% CI, -1.5 to 3.0 mph).

Bland-Altman plots indicating the mean differences in GXT VT1 and VT2 heart rate between TRUE and S-PRED and levels of agreement with 95% CIs are illustrated in Figure 7 and Figure 8. For trial 1 VT1 heart rate, the mean difference between TRUE and S-PRED was 4.0 ± 7.2 bpm (95% CI, -10.1 to 18.1 bpm). For trial 2 VT1 heart rate, the mean difference between TRUE and S-PRED was 4.3 ± 11.1 bpm (95% CI, -17.6 to 26.1 bpm). For trial 1 VT2 heart rate, the mean difference between TRUE and S-PRED was 5.3 ± 8.6 bpm (95% CI, -11.7 to 22.2 bpm). For trial 2 VT2 heart rate, the mean difference between TRUE and S-PRED was 8.2 ± 13.7 bpm (95% CI, -18.7 to 35.1 bpm).

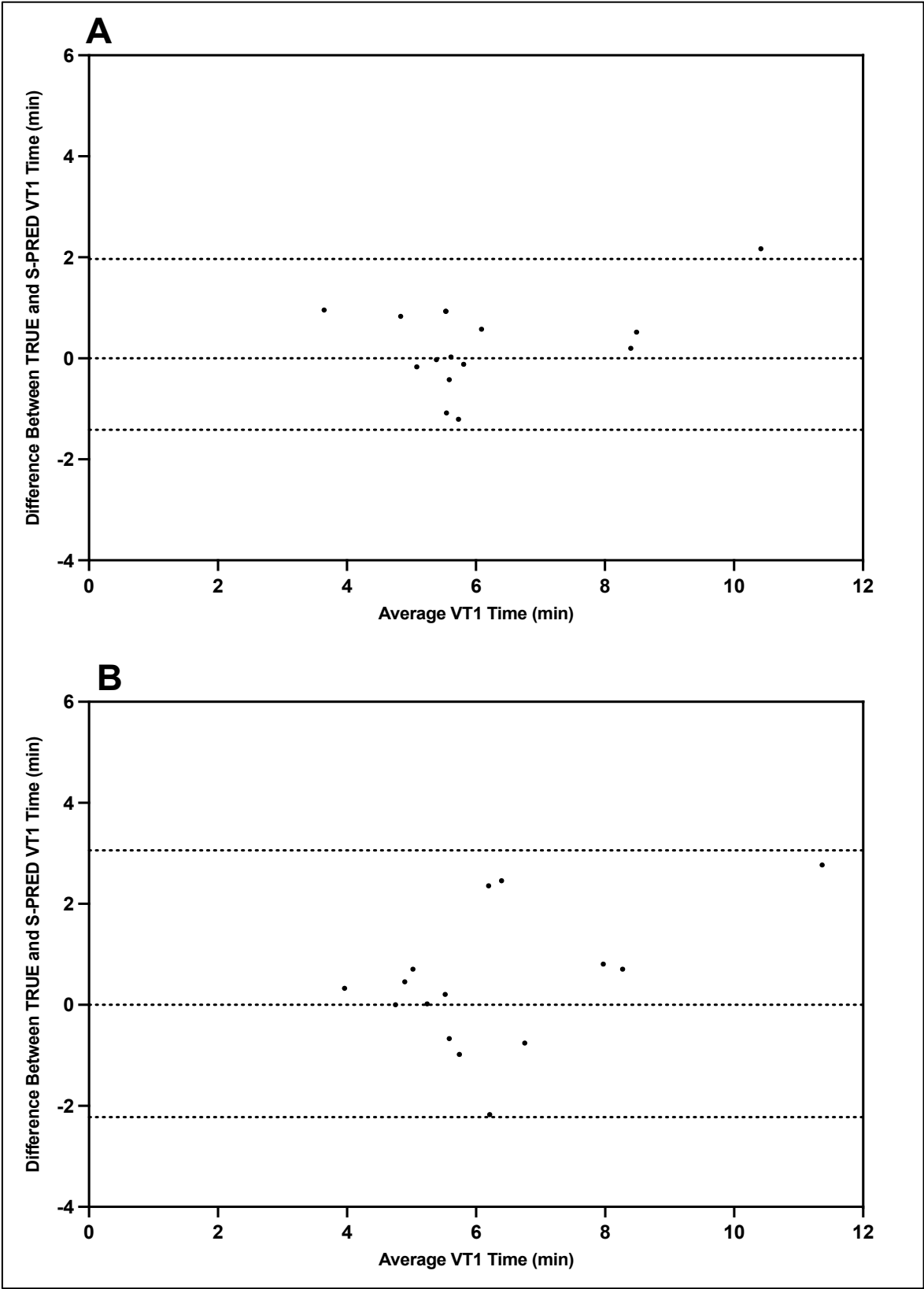


Figure 3. TRUE versus S-PRED VT1 time point; trial 1 (A) and trial 2 (B).

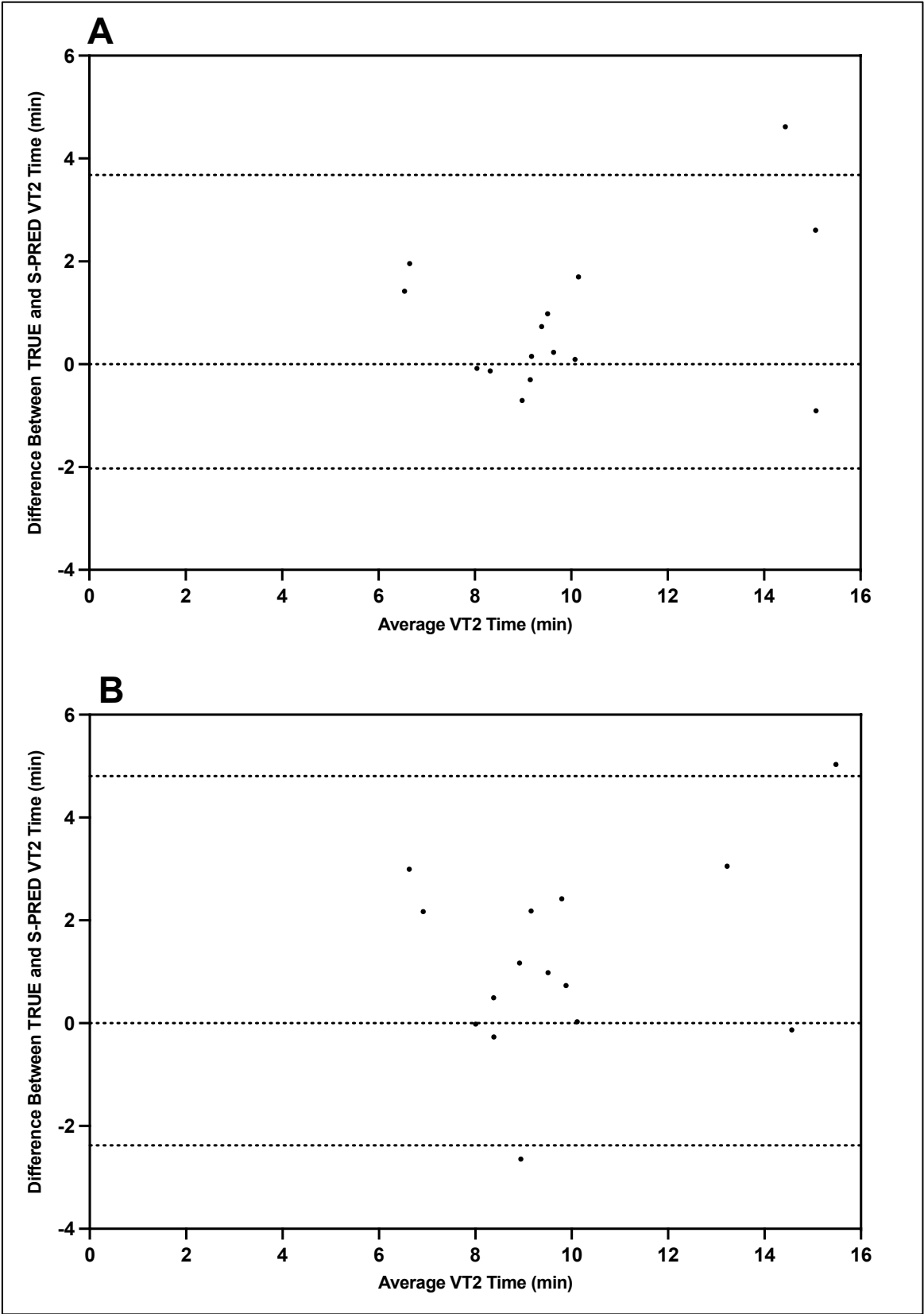


Figure 4. TRUE versus S-PRED VT2 time point; trial 1 (A) and trial 2 (B).

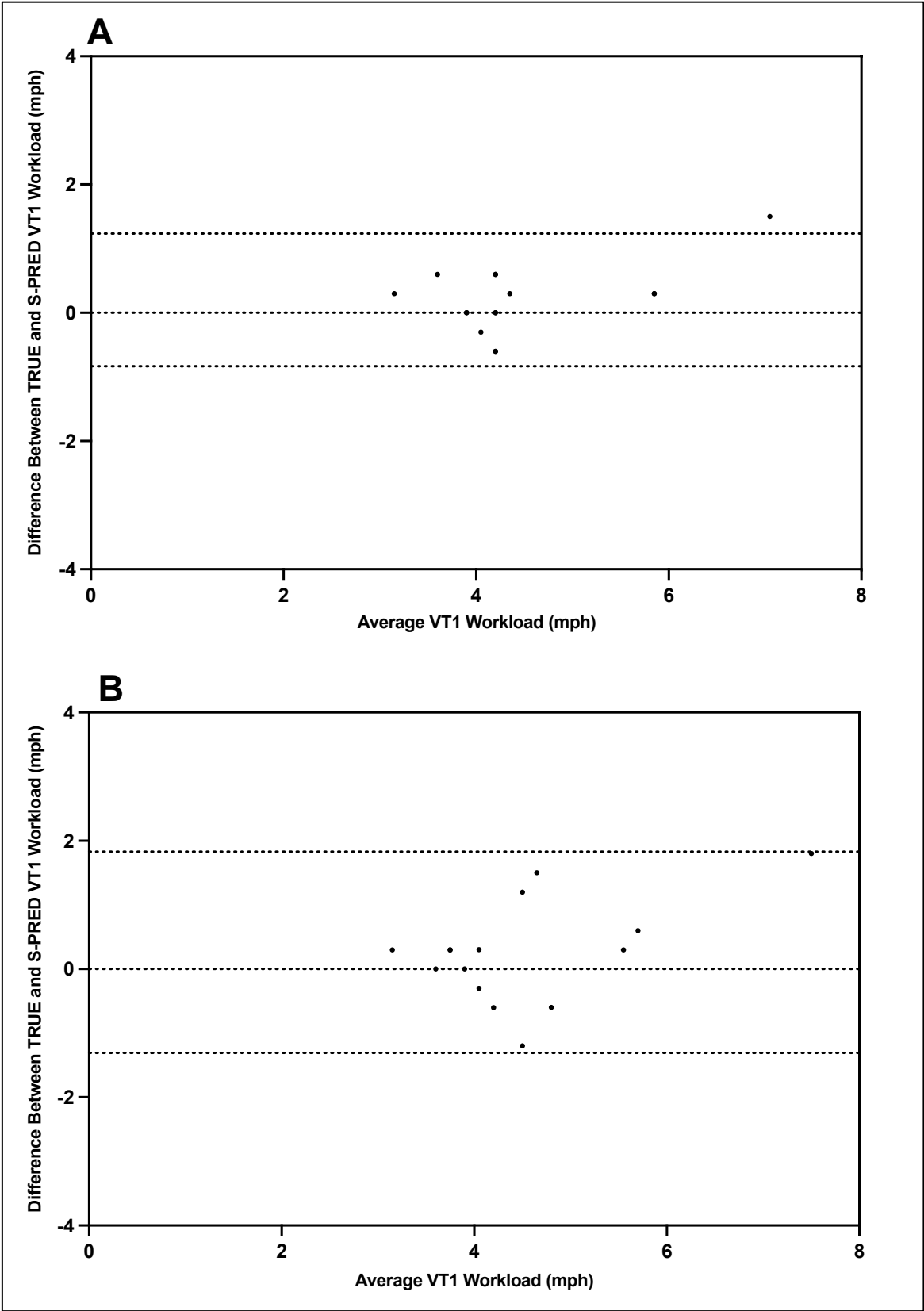


Figure 5. TRUE versus S-PRED VT1 workload; trial 1 (A) and trial 2 (B).

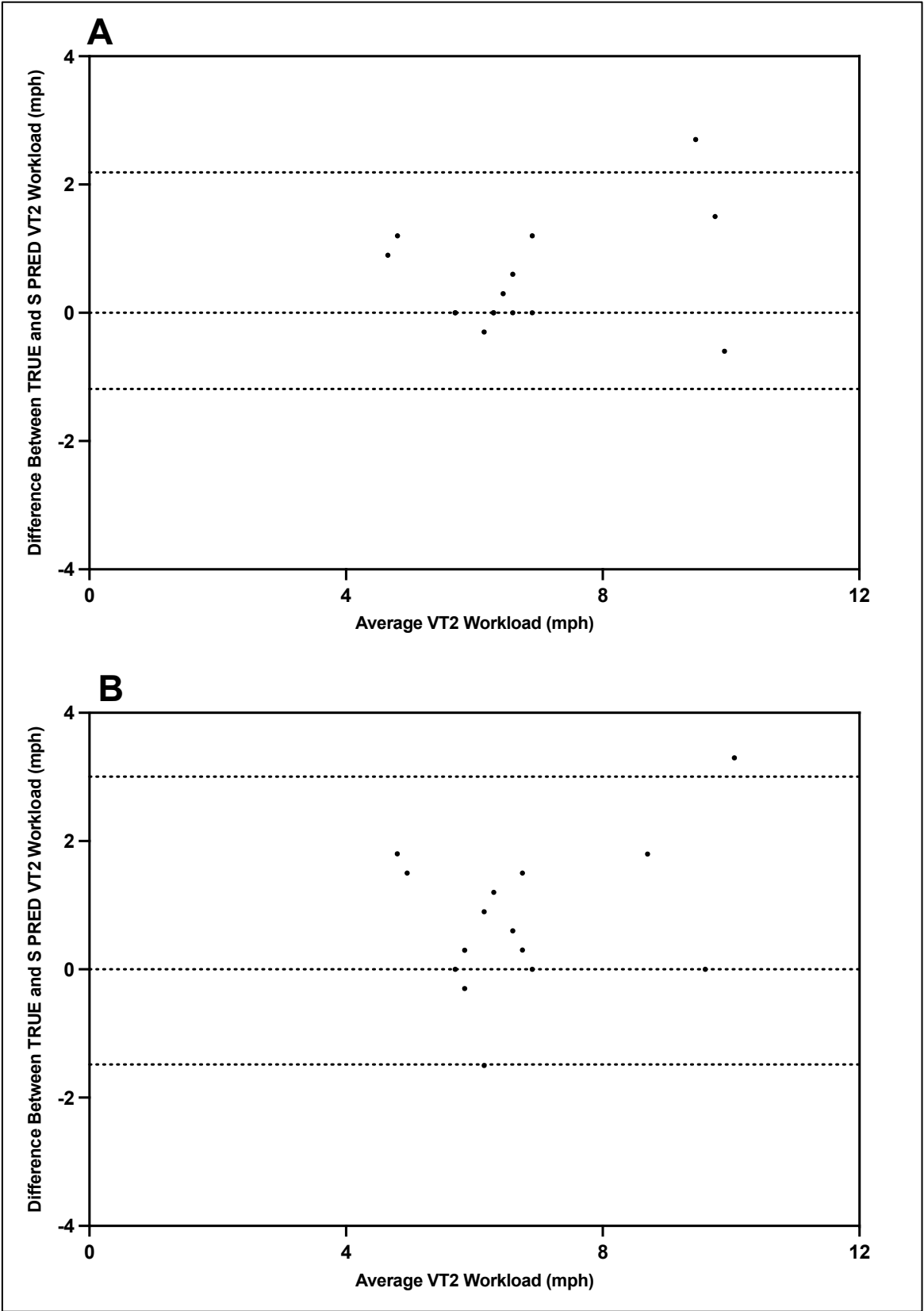


Figure 6. TRUE versus S-PRED VT2 workload; trial 1 (A) and trial 2 (B).

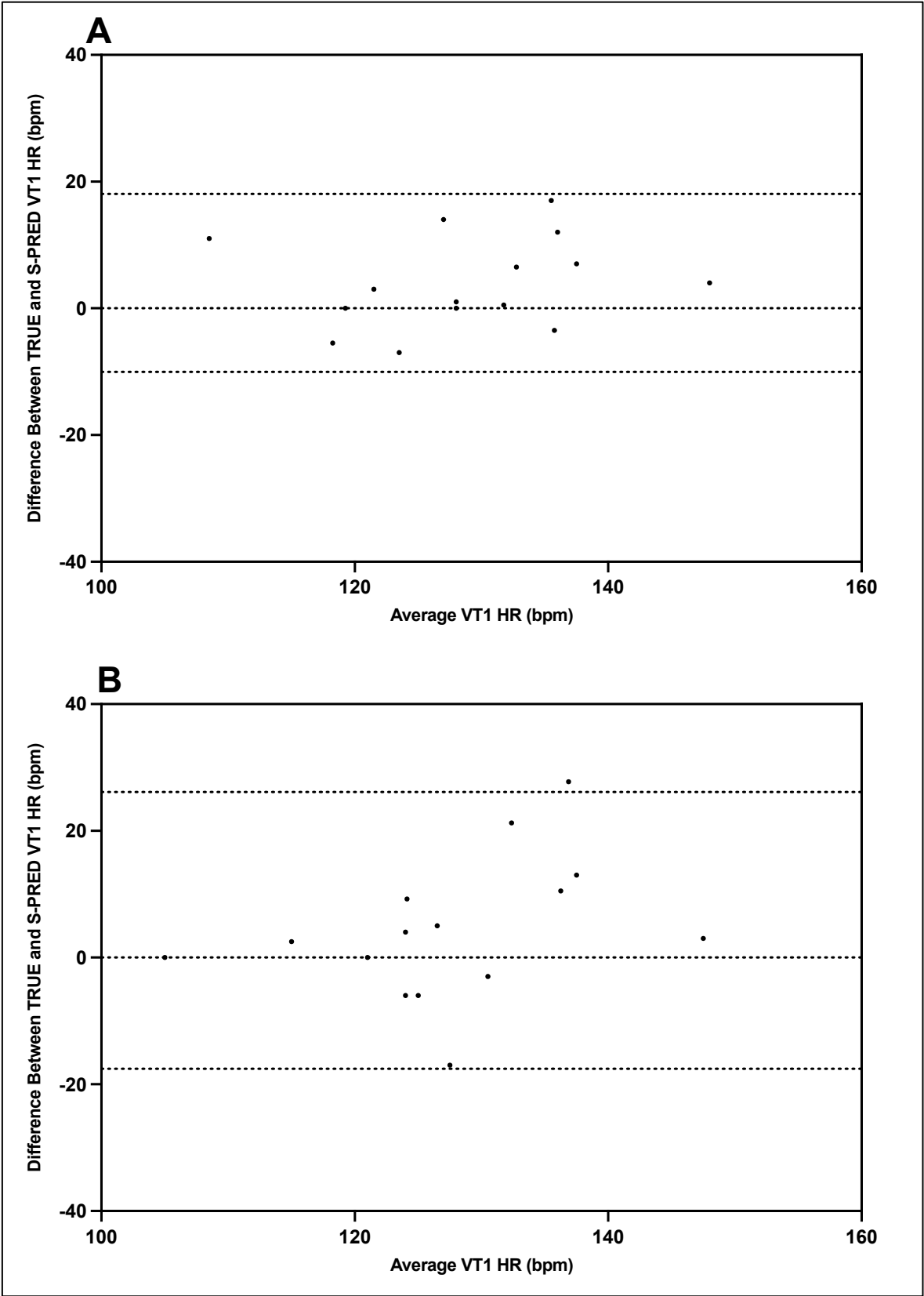


Figure 7. TRUE versus S-PRED VT1 heart rate; trial 1 (A) and trial 2 (B).

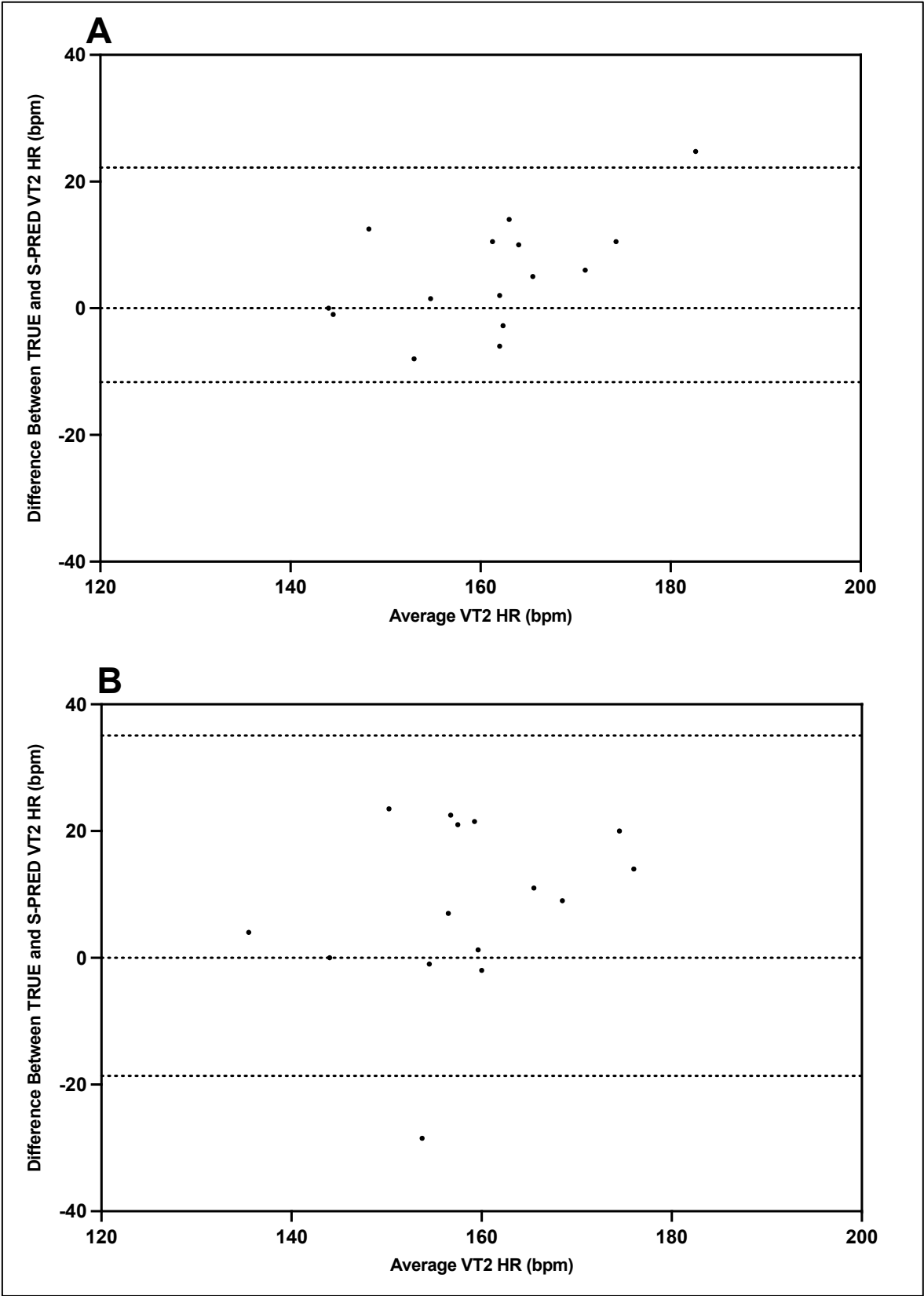


Figure 8. TRUE versus S-PRED VT2 heart rate; trial 1 (A) and trial 2 (B).

4. Discussion

4.1 Primary findings

The aim of this study was to evaluate the reliability and validity of the Tyme Wear smart shirt to predict VT1 and VT2 when compared to the gold standard determined by open-circuit indirect calorimetry. Based on the findings from this study, the ventilatory threshold data suggests that the Tyme Wear smart shirt (S-PRED) is similar in reliability when compared to that of the gold standard (TRUE). However, the ventilatory threshold data suggests that S-PRED is less valid relative to TRUE. Consequently, there is a tendency for S-PRED to underestimate the workload and heart rate values corresponding to VT1 and VT2 compared to TRUE VT1 and VT2. These novel and encouraging findings provide the necessary data to establish the foundation to determine the efficacy of wearable technologies to guide threshold-based training.

4.2 Reliability – cohort level

Data from the current study supports previous findings of TRUE's *excellent* reliability with ICC values greater than 0.9 across all trials for timepoint, heart rate, and workload corresponding to VT1 and VT2 [15-17]. Similarly, S-PRED also produced *excellent* reliability utilizing time point to determine VT1 and VT2; *moderate* reliability in determining VT1 and VT2 heart rate; *excellent* reliability when using workload to determine VT1 and VT2.

Another important finding from this study are the expected technical error values, a combination of both biological error and measurement error, for VT1 and VT2. For example, TRUE data suggests the technical error, as CV, for VT1 and VT2 workload to be 6.8% and 4.9%, respectively. Furthermore, TRUE data suggests that the technical error, as CV, for heart rate at VT1 and VT2 to be 3.2% and 2.4%, respectively. The results in the current study for CV are within similar ranges to those found elsewhere where CV for VT1 workload and heart rate were 5.2% and 4.3%, respectively, while CV for VT2 workload and heart rate were 2.4% and 4.2%, respectively [18]. The finding from the current study for CV VT1 workload is also similar to that suggested by Dickhurth and colleagues (CV = 5.8%) [19]. Interestingly, in a study by Gasparini Neto and colleagues, the suggested heart rate CV values for non-athletes at VT1 and VT2 were 11.8% and 6.6%, respectively while elite athletes heart rate CV values for VT1 and VT2 were 3.5% and 5.1%, respectively [20]. The authors suggest that the higher CV in the non-athletes could have been due to large inter-stage increases in intensity during the GXT which may not be appropriate for this population. As such, it appears the differences in published CV values can be attributed to the GXT protocol, the population under study, equipment used to measure gas exchange, and methods used to detect VT1 and VT2 [21].

4.3 Validity – cohort level

The data presented in Table 4 suggests that S-PRED underestimates the timepoint, workload, and heart rate values corresponding to VT1 and VT2 when compared to TRUE. Furthermore, the bias and wide limits of agreement revealed in the Bland Altman plots suggest that the agreement between TRUE and S-PRED is less than optimal for VT1 and even less so for VT2. The difference in workload is approximately 2-3% and 5-9% between TRUE and S-PRED VT1 and VT2, respectively, as observed in Table 6. For heart rate, the difference is approximately 2% and 2-5% between TRUE and S-PRED VT1 and VT2, respectively, as observed in Table 5. This translates to a workload difference of 0.2-0.3 mph and 0.5-0.9 mph between TRUE and S-PRED VT1 and VT2, respectively. For heart rate, this translates to a difference of 3.7 BPM and 3.7-9.2 BPM between TRUE and S-PRED VT1 and VT2, respectively. As such, it appears that for the cohort in our study, for every 0.3 mph increase, heart rate increases concomitantly by 5-10 BPM.

However, is the underestimation of S-PRED heart rate and workload outside the physiological range for VT1 and VT2? A study by Gasparini Neto et al. [20], suggested that for non-athletes, VT1

occurred at $76.0 \pm 5.4\%$ of heart rate max while elite athletes VT1 occurred at $85.5 \pm 4.2\%$. Their data also suggests that VT2 for non-athletes occur at $91.4 \pm 4.0\%$ of heart rate max, while VT2 for elite athletes occurred at $94.9 \pm 2.4\%$ of heart rate max [20]. In persons with chronic stroke, Boyne et al. [6] suggested that VT1 occurred at 80% of peak heart rate, while the ranges for effectively exercising at VT1 ranged from 55% to 80% of peak heart rate. In another study, Cerezuela-Espejo and colleagues suggested that the 95% confidence interval for VT1 and VT2 heart rate were 77-81% and 91-93% of max heart rate, respectively and that VT1 and VT2 workload were 59-65% and 84-87% of maximum achieved speed, respectively [22]. Taken together, the ranges for VT1 and VT2 suggested in the aforementioned studies suggests that the underestimation observed in Tables 5 and 6 for S-PRED VT1 may not be physiologically significant, but the underestimation for S-PRED VT2 may be physiologically significant. As such, it can be postulated that for our cohort, VT1 heart rate can range within 15-20 BPM and remain within the exercise intensity range of VT1 while VT2 heart rate range can be within 12 BPM and remain within the exercise intensity range of VT2.

4.4 S-PRED validity – individual level

While accounting for the individual variations that fall beyond the aforementioned ranges for VT1 and VT2 workload and heart rate, which can be observed in the Bland Altman plots, S-PRED may be effective identifying workloads corresponding to VT1 and VT2 for approximately 83% and 57% of the cohort and similar populations, respectively. Additionally, S-PRED may be effective at detecting heart rate values related to VT1 and VT2 for approximately 86% and 67% of this cohort, respectively. However, the underestimation appears to be exacerbated for a select number of subjects in this cohort. Specifically, these underestimations for S-PRED heart rate were exacerbated in one and three individuals for VT1 and VT2, respectively, in trial 1. The ranges of the severe underestimations in trial 1 were 8.4% and 7.1-12.3% below TRUE for VT1 and VT2, respectively. Furthermore, these underestimations for S-PRED heart rate were exacerbated in three and seven individuals for VT1 and VT2, respectively, in trial 2. The ranges of these more severe underestimations in trial 2 were 8.8-15.0% and 7.21-14.7% below TRUE for VT1 and VT2, respectively. In terms of workload, the underestimations were exacerbated in one and five individuals for VT1 and VT2, respectively, for trial 1. The ranges of these more severe underestimations in trial 1 were 11.6% and 11.6-22.0% below TRUE for VT1 and VT2, respectively. Additionally, in trial 2, the underestimations were exacerbated in four and eight individuals for the workloads corresponding to VT1 and VT2, respectively. The ranges of these more severe underestimations in trial 2 were 12.5-16.1% and 10.3-27.3% below TRUE for VT1 and VT2, respectively.

Upon closer inspection of the Bland Altman plots, there are four particular subjects whose underestimation are consistently larger than the others for workload and heartrate at VT1 and VT2 across both trials 1 and 2. Specifically, in Figure 6 and Figure7, it appears that these four subjects that had the greatest difference between S-PRED and TRUE also had higher fitness levels ($\text{VO}_{2\text{max}} > 60 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for males and $> 50 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for females) when compared to the rest of the cohort, whereby VT1 and VT2 workload occur at higher intensities. One possible explanation for this is that the algorithms used to model VT1 and VT2 are not well suited for those users with greater fitness levels and may be unable to accurately detect VT1 and VT2 workloads for these types of individual.

4.5 Potential explanation for reduced validity of the Tyme Wear smart shirt

There are several reasons why S-PRED is less valid when compared to the TRUE. There are two points where error can arise: 1) data collected by the sensors; 2) the accuracy of the algorithms to correctly detect ventilatory thresholds. In regard to the sensor data collection, S-PRED utilizes proxy measurements (sensor stretch and rate of stretch) to replicate the equivalent data (tidal volume and respiratory rate, respectively). Furthermore, we are unaware of the inter-unit measurement error of the sensors. Additionally, we are unfamiliar with the algorithms used to determine VT1 and VT2. Without the ability to compare S-PRED VE and TRUE VE in liters per minute, it is difficult to conduct

an equivalent comparison and to determine whether the S-PRED underestimations of VT1 and VT2 are due to the sensors or algorithms. If the underestimation of VE is the primary cause of the underestimation of VT1 and VT2 for S-PRED, it is mostly likely that the sensor is underestimating tidal volume. Based on the understanding of the technology, respiratory rate would be easily determined, but the estimations for tidal volume may be more difficult as the fit of the shirt and the sensor's sensitivity to the expansion of the thoracic cavity during a respiratory cycle creates potential for measurement error. However, it is plausible that sensor and algorithm error are contributing to the underestimation.

Finally, the cost of the TrueOne2400 is magnitudes greater in cost than the Tyme Wear smart shirt. Simply put, there must be an expectation where the smart shirt will not be as accurate or reliable as the TrueOne2400. However, this does not insinuate that the Tyme Wear smart shirt is ineffective at estimating VT1 and VT2. Rather, it appears that the sensors, algorithms, or both may need further refinement to improve capture accuracy and ventilation data processing to establish greater reliability and validity when detecting ventilatory thresholds for a segment of the population with high levels of fitness.

4.6 Value and efficacy of the Tyme Wear smart shirt to detect ventilatory thresholds

The goal of the Tyme Wear smart shirt is improve performance by optimizing training with the use of VT1 and VT2. Using ventilatory thresholds to guide training has been shown to improve training responsiveness when compared to standardized training. In this current study, the primary concern with the underestimation of the S-PRED is that the suggested training intensity for VT1 and VT2 may not be sufficient to elicit the necessary and desired adaptive response to prescribed exercise and workouts. Even though the shirt may underestimate the exercise intensities, it can be argued that personalized, structured training within specified, albeit underestimated for S-PRED, workload zones is superior to standardized training [5, 11-12, 23]. Thus, the question arises: Is the underestimation by S-PRED significant enough to detrimentally impact training and performance? While a future study should address this question, it is possible to speculate on the outcomes. The consequence of training below the true physiological VT1 and VT2 means that users may not receive the adequate dose of exercise stress needed to elicit improvements in exercise performance. However, depending on the individual utilizing the smart shirt, an untrained individual may observe improvements in their exercise performance simply due to the introduction of exercise stress. Contrastingly, an elite athlete, for example, may need more precision than the smart shirt offers, specifically for VT2, as sufficient exercise stress at specific intensities is required to improve performance in a well-trained system.

4.7 Novelty of study and additional findings

To our knowledge, this is the first study to examine the reliability and validity of a wearable technology that measures respiratory rate and tidal volume to calculate VE in order to detect VT1 and VT2. Traditionally, ventilatory thresholds are determined through gas exchange data as functions of $\dot{V}O_2$, $\dot{V}CO_2$, and VE, whereby plots of $\dot{V}E/\dot{V}O_2$ is used to determine VT1 and plots of $\dot{V}E/\dot{V}CO_2$ is used to determine VT2 [24]. Examples of these methods include the ventilation curve method, V-slope method, and ventilatory equivalents method [8]. Furthermore, there are field based methods such as the talk test which can determine VT1 with high reliability and validity [25-27]. However, our lab recently conducted a retrospective analysis and found that the sole use of VE was a sufficient and viable way to estimate VT1 and VT2 [9]. This finding supports the paradigm that the Tyme Wear smart shirt has the potential to accurately detect VT1 and VT2 by exclusively using VE.

Additionally, our study provides evidence to support the ranges for percent of maximum heart rate and workload at VT1 and VT2. Our data suggests that the percent of peak heart rate for VT1 is approximately 70% with a range of 60-80%, while VT2 is approximately 88% with a range of 80-95%.

Additionally, our data suggests that the percent of maximum workload for VT1 is approximately 50% with a range of 40-60%, while VT2 is approximately 75% with a range of 60-90%. Table 7 compares the findings for ranges of this study with ranges from other studies. The differences in previously reported VT1 and VT2 ranges compared to the ranges found in the current study may be explained by the heterogenous fitness of the subjects in the current study; a homogenous fitness population would create smaller ranges for heart rate and workload VT1 and VT2, while a heterogenous population would create larger ranges. Furthermore, previous studies have found that with exercise training, VT1 and VT2 occurs at greater percentages of VO_2max [28-30]. This suggests that individuals with lower fitness, defined by VO_2max , achieve VT1 and VT2 at lower percentages of their maximal heart rate and workload intensities. However, in recovering stroke patients, heart rate at the ventilatory threshold was predicted and found to be approximately 80% of peak heart rate in a symptom limited GXT, suggesting illness can alter this relationship [6]. As such, these factors can account for the differences in reported ranges for workload and heart rate at VT1 and VT2.

Table 7. Comparison of VT1 and VT2 workload and heart rate ranges.

	Data From TRUE	Cerezuela-Espejo et al.	Boyne et al.	Gasparini Neto et al.
Population Type	Recreationally Active	Trained Athletes	Chronic Stroke Patients	Non-Trained Elite Athletes
VO_2max or peak ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$)	48.4 ± 10.5	60.2 ± 4.3	16.2 ± 5.7	47.2 ± 4.4 68.6 ± 3.2
Workload VT1 (% of Max Speed)	48.5 ± 6.5 (Range: 39 – 62)	59-65	-	57.1 ± 6.9 70.0 ± 6.8
Workload VT2 (% of Max Speed)	74.5 ± 8.0 (Range: 59.5 – 88.5)	84-87	-	80.4 ± 9.0 87.7 ± 4.1
Heart Rate VT1 (% of Max or Peak Heart Rate)	70.5 ± 6.0 (Range: 60 – 83)	77-81	80.0 ± 8.0 (Range: 68-94)	76 ± 5.4 85.5 ± 4.2
Heart Rate VT2 (% of Max or Peak Heart Rate)	88 ± 6.0 (Range: 75 – 97.5)	91-93	-	91.4 ± 4.0 94.9 ± 2.4

Values are mean \pm SD.

4.8 Limitations

In this study, there are several limitations that should be mentioned. This includes requesting, but not controlling for subjects' pre and post testing lifestyle behaviors. Not controlling for these extraneous lifestyle behaviors and variables such as diet, sleep, and prior exercise training could have contributed to the increased ranges for VT1 and VT2. Furthermore, we did not control for time of day that testing took place. Previous studies had found that circadian rhythm could affect maximal aerobic power and time to exhaustion [31-32]. Additionally, we are unaware of the inter-unit

variability for measurement error. As such, the validity data could be affected by the different smart shirts used. Finally, the heterogeneity of subjects' fitness levels may have contributed to the wider ranges for VT1 and VT2 found in the current study compared to the ranges in literature. With this in mind, further studies may look to examine the efficacy of the smart shirt to detect VT1 and VT2 in cohorts that have homogenous fitness.

5. Conclusions

The Tyme Wear smart shirt allow the opportunity for users to have relatively affordable access to information that was once only obtainable through a laboratory with the appropriate equipment and typically more expensive cost of testing. However, the caveat is that the detection of VT1 and VT2 by the Tyme Wear smart shirt, at its current stage in development, is less valid with similar reliability to its laboratory counterpart. However, the suggested ranges for exercise intensities for VT1 and VT2 should be an adequate workload for most individuals, although caution should be taken with individuals with high fitness. As such, the Tyme Wear smart shirt provides easily accessible testing to establish threshold-guided training zones but will unlikely replace or devalue the need for the laboratory equivalent.

Author Contributions: Conceptualization, A.G., G.V., and L.D.; Methodology, A.G., G.V., G.C., M.K., and L.D.; Formal Analysis, A.G., G.V., and L.D.; Investigation, A.G., G.V., G.C., M.K., and L.D.; Resources, L.D.; Data Curation, A.G. and L.D.; Writing—Original Draft Preparation, A.G., G.V., and L.D.; Writing—Review and Editing, A.G., G.V., G.C., M.K., and L.D.; Visualization, L.D.; Supervision, G.V. and L.D.; Project Administration, L.D.; Funding Acquisition, L.D.

Funding: The American Council on Exercise funded this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Grand View Research. Wearable Technology Market Growth & Trends. Available online: <https://www.grandviewresearch.com/industry-analysis/wearable-technology-market> (accessed on 31 October 2020).
2. Reid, C.R.; Schall, M.C.; Amick, R.Z.; Schiffman, J.M.; Lu, M.L.; Smets, M.; Moses, H.R.; Porto, R. Wearable technologies: how will we overcome barriers to enhance worker performance, health, and safety? *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2017**, *61*, 1026–1030. <https://doi.org/10.1177/1541931213601740>
3. Jhangiani, R.; Chiang, I.-C.; Cuttler, C.; Leighton, D. (2019). *Research Methods in Psychology*, 2019. <https://doi.org/10.17605/OSF.IO/HF7DQ>
4. Wahl, Y.; Düking, P.; Droszez, A.; Wahl, P.; & Mester, J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. *Front. Physiol.* **2017**, *8*. <https://doi.org/10.3389/fphys.2017.00725>
5. Astorino, T.A.; deRevere, J.; Anderson, T.; Kellogg, E.; Holstrom, P.; Ring, S.; Ghaseb, N. Change in VO₂max and time trial performance in response to high-intensity interval training prescribed using ventilatory threshold. *Eur. J. Appl. Physiol.* **2018**, *118*, 1811–1820. <https://doi.org/10.1007/s00421-018-3910-3>
6. Boyne, P.; Buhr, S.; Rockwell, B.; Khoury, J.; Carl, D.; Gerson, M.; Kissela, B.; Dunning, K. Predicting heart rate at the ventilatory threshold for aerobic exercise prescription in persons with chronic stroke. *J. Neurol. Phys. Ther.*, **2015**, *39*, 233–240. <https://doi.org/10.1097/NPT.0000000000000102>

7. Emerenziani, G.P.; Ferrari, D.; Marocco, C.; Greco, E. A.; Migliaccio, S.; Lenzi, A.; Baldari, C.; Guidetti, L. Relationship between individual ventilatory threshold and maximal fat oxidation (MFO) over different obesity classes in women. *PLoS One*, **2019**, 14(4):e0215307. <https://doi.org/10.1371/journal.pone.0215307>
8. Gaskill, S.E.; Ruby, B.C.; Walker, A.J.; Sanchez, O.A.; Serfass, R.C.; Leon, A.S. Validity and reliability of combining three methods to determine ventilatory threshold: *Med. Sci. Sports Exerc.*, **2001**; 33, 1841–1848. <https://doi.org/10.1097/00005768-200111000-00007>
9. Gouw, A.H.; Van Guilder, G.P.; Larusson, A.; Laredo, G.; Weatherwax, R.M.; Byrd, B.R.; Dalleck, L.C. Ventilation can exclusively be used to predict ventilatory thresholds: a retrospective analysis. *Int. J. Res. Ex. Phys.* **2021**, 16, 1–18.
10. Howley, E.T.; Bassett, D.R. Jr; Welch, H.G. Criteria for maximal oxygen uptake: review and commentary. *Med. Sci. Sports Exerc.* **1995**, 27, 1292–1301.
11. Wolpern, A.E.; Burgos, D.J.; Janot, J.M.; Dalleck, L.C. Is a threshold-based model a superior method to the relative percent concept for establishing individual exercise intensity? A randomized controlled trial. *BMC Sports Sci. Med. Rehabil.* **2015**, 7, 16. <https://doi.org/10.1186/s13102-015-0011-z>
12. Weatherwax, R.M.; Harris, N.K.; Kilding, A.E.; Dalleck, L. C. Incidence of vo2max responders to personalized versus standardized exercise prescription. *Med. Sci. Sports Exerc.* **2019**, 51, 681–691. <https://doi.org/10.1249/MSS.0000000000001842>
13. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
14. Bland, J.M., Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, 1(8476), 307–310.
15. Stuart, M.K.; Howley, E.T.; Gladden, L.B.; Cox, R.H. Efficiency of trained subjects differing in maximal oxygen uptake and type of training. *J. Appl. Physiol. Respir. Environ. Exerc. Physiol.* **1981**, 50, 444–449. <https://doi.org/10.1152/jappl.1981.50.2.444>
16. Macfarlane, D.J.; Wu, H.L. Inter-unit variability in two ParvoMedics TrueOne 2400 automated metabolic gas analysis systems. *Eur. J. Appl. Physiol.* **2013**, 113, 753–762. <https://doi.org/10.1007/s00421-012-2483-9>
17. Crouter, S.E.; Antczak, A.; Hudak, J.R.; DellaValle, D.M.; Haas, J.D. Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. *Eur. J. Appl. Physiol.* **2006**, 98, 139–151. <https://doi.org/10.1007/s00421-006-0255-0>
18. Lourenço, T.F.; Martins, L.E.B.; Tessutti, L.S.; Brenzikofer, R.; Macedo, D.V. Reproducibility of an incremental treadmill vo2max test with gas exchange analysis for runners. *J. Strength Cond. Res.* **2011**, 25, 1994–1999. <https://doi.org/10.1519/JSC.0b013e3181e501d6>
19. Dickhuth, H.H.; Yin, L.; Niess, A.; Röcker, K.; Mayer, F.; Heitkamp, H.C.; Horstmann, T. Ventilatory, lactate-derived and catecholamine thresholds during incremental treadmill running: relationship and reproducibility. *Int. J Sports Med.* **1999**; 20, 122–127. <https://doi.org/10.1055/s-2007-971105>
20. Gasparini Neto, V.H.; Azevedo, P.; Carletti, L.; Perez, A.J. Oxygen consumption and heart rate obtained in a ramp protocol are equivalent during exercise session of rectangular loading at ventilatory thresholds for athletes. *Rev. bras. cineantropom. desempenho hum.* **2019**, 21, e58285. <https://doi.org/10.1590/1980-0037.2019v21e58285>
21. Meyer, K.; Hajric, R.; Westbrook, S.; Samek, L.; Lehmann, M.; Schwaibold, M.; Betz, P.; Roskamm, H. Ventilatory and lactate threshold determinations in healthy, normal and cardiac patients: Methodological problems. *Eur. J. Appl. Physiol. Occup. Physiol.* **1996**, 72, 387–393. <https://doi.org/10.1007/BF00242266>

22. Cerezuela-Espejo, V.; Courel-Ibáñez, J.; Morán-Navarro, R.; Martínez-Cava, A.; Pallarés, J.G. The relationship between lactate and ventilatory thresholds in runners: validity and reliability of exercise test performance parameters. *Front. Physiol.* **2018**, *9*, 1320. <https://doi.org/10.3389/fphys.2018.01320>
23. Fabre, C.; Massé-Biron, J.; Ahmaidi, S.; Adam, B.; Préfaut, C. Effectiveness of individualized aerobic training at the ventilatory threshold in the elderly. *J. Gerontol. A Biol. Sci. Med. Sci.* **1997**, *52*, B260–B266. <https://doi.org/10.1093/gerona/52A.5.B260>
24. Wasserman, K.; McIlroy, M.B. Detecting the threshold of anaerobic metabolism in cardiac patients during exercise. *Am. J. Cardiol.* **1964**, *14*, 844–852. [https://doi.org/10.1016/0002-9149\(64\)90012-8](https://doi.org/10.1016/0002-9149(64)90012-8)
25. Ballweg, J.; Foster, C.; Porcari, J.; Haible, S.; Aminaka, N.; Mikat, R.P. Reliability of the talk test as a surrogate of ventilatory and respiratory compensation thresholds. *J. Sports Sci. Med.* **2013**, *12*, 610–611.
26. Rodríguez-Marroyo, J.A.; Villa, J.G.; García-López, J., Foster, C. Relationship between the talk test and ventilatory thresholds in well-trained cyclists. *J. Strength Cond. Res.* **2013**, *27*, 1942–1949. <https://doi.org/10.1519/JSC.0b013e3182736af3>
27. Zanettini, R.; Centeleghe, P.; Franzelli, C.; Mori, I.; Benna, S.; Penati, C.; Sorlini, N. Validity of the talk test for exercise prescription after myocardial revascularization. *Eur. J. Prev. Cardiol.* **2013**, *20*, 376–382. <https://doi.org/10.1177/2047487312438982>
28. Röcker, K.; Striegel, H.; Freund, T.; Dickhuth, H.H. Relative functional buffering capacity in 400-meter runners, long-distance runners and untrained individuals. *Eur. J. Appl. Physiol. Occup. Physiol.* **1994**, *68*, 430–434. <https://doi.org/10.1007/BF00843741>
29. Esteve-Lanao, J.; Foster, C.; Seiler, S.; Lucia, A. Impact of training intensity distribution on performance in endurance athletes. *J. Strength Cond. Res.* **2007**, *21*, 943–949. <https://doi.org/10.1519/R-19725.1>
30. Zapico, A.G.; Calderón, F.J.; Benito, P.J.; González, C.B.; Parisi, A.; Pigozzi, F.; Di Salvo, V. Evolution of physiological and hematological parameters with training load in elite male road cyclists: a longitudinal study. *J. Sports Med. Phys. Fitness* **2007**, *47*, 191–196.
31. Hill, D.W.; Cureton, K.J.; Collins, M.A. Effect of time of day on perceived exertion at work rates above and below the ventilatory threshold. *Res. Q. Exerc. Sport* **1989**, *60*, 127–133. <https://doi.org/10.1080/02701367.1989.10607427>
32. Carter, H.; Jones, A.M.; Maxwell, N.S., Doust, J.H. The effect of interdiurnal and diurnal variation on oxygen uptake kinetics during treadmill running. *J. Sports Sci.* **2002**, *20*, 901–909. <https://doi.org/10.1080/026404102320761796>