Open Access Journal of Sports Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Garmin Fénix 7® Underestimates Performance at the Lactate Threshold in Comparison to Standardized Blood Lactate Field Test

Marie Heiber[1,*], Andrea Schittenhelm[1,*], Jennifer Schlie[2,*], Marcus Beckert[2], Pascal Graf[2], Annette Schmidt[1–3]

[1]dtec.bw, NextGenerationEU Project Smart Health Lab, University of the Bundeswehr, Chair of Sport Biology Munich, Munich, Germany; [2]University of the Bundeswehr, Institute of Sport Sciences, Chair of Sport Biology, Munich, Germany; [3]Research Center Smart Digital Health, University of the Bundeswehr Munich, Munich, Germany

*These authors contributed equally to this work

Correspondence: Annette Schmidt, Chair of Sport Biology, University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, Neubiberg, Munich, Bavaria, 85577, Germany, Tel +49 89 6004 4412, Email annette.schmidt@unibw.de

Purpose: Lactate threshold (LT) is a critical performance measure traditionally obtained using costly laboratory-based tests. Wearables offer a practical and noninvasive alternative for LT assessment in recreational and professional athletes. However, the comparability of these estimates with the regular field tests requires further evaluation.

Patients and Methods: In our sample of 26 participants (n_f=7 and n_m=19), we compared the estimated running pace and heart rate (HR) at LT with two subsequent tests. First, participants performed the Fenix 7® threshold running test after a calibration phase. Subsequently, they were tested in a standardized, graded blood lactate field test. Age was 25.97 (± 6.26) years, and body mass index (BMI) was 24.58 (± 2.8) kg/m^2.

Results: Pace at LT calculated by Fenix 7® (*M*=11.87 km/h ± 1.26 km/h) was 11.96% lower compared to the field test (*M*=13.28 km/h ± 1.72 km/h), which was significant (p <0.001, d=−1.19). HR estimated by the Fenix 7® at LT was 1.71% lower (p >0.05). LT data obtained in the field test showed greater overall variance.

Conclusion: Our results suggest sufficient accuracy of Fenix 7® LT estimates for recreational athletes. It can be assumed that for professional athletes, it would fail to provide the nuanced data needed for high-quality training management.

Keywords: smartwatch, physical performance, physiology, heart rate, pace

Introduction

Performance diagnostics are a central element for many professional and recreational athletes throughout the season. These diagnostics usually apply advanced technologies such as calorimetry, photoplethysmography, near-infrared-spectroscopy (NIRS), blood lactate analysis, and ergospirometry. Among the numerous measures, cardiorespiratory fitness (VO_{2max}) and lactate threshold (LT) are crucial performance indicators, particularly in endurance disciplines.[1,2] Previously, athletes depended on time-consuming and costly laboratory-based tests to obtain these metrics. The gold standard procedures for ergospirometry for VO_{2max} and lactate testing for LT[3] involve wearing uncomfortable masks and invasive blood sample acquisition. After testing, the analysis and implementation of the acquired data for training and racing were reserved for sports scientists and coaches. In recent years, different non-invasive methods for LT determination that use surrogate markers and algorithms to predict lactate levels have been introduced.[4] One example is the use of wearable NIRS devices attached to prominent locomotor muscles during a given exercise. During running, NIRS can be used to track changes in muscle oxygenation in the calf or quadriceps muscles in response to running speed.[5] Despite the non-invasiveness of these methods and evidence of their validity and reliability in detecting LT,[5] their application is still costly and rather laboratory-based. It requires trained personnel for the measurement and interpretation. With the advent of wearables such as global positioning

Received: 20 October 2023
Accepted: 7 March 2024
Published: 9 May 2024

system (GPS)-coupled fitness watches, performance testing should become accessible to a broad range of customers. These smartwatches aim to master the advanced features of performance testing, such as VO_{2max} and LT determination, and provide training and recovery recommendations, as well as race result predictions.

Among the general population, smartwatches are popular for their daily activity tracking with measures like step count, climbed floors, heart rate (HR), and, based on that, an estimate of daily energy expenditure. However, smartwatches play a more profound role for recreational and professional athletes. The features of advanced models have evolved to include complex physiological measures such as heart rate variability, $VO_{2max,}$ and LT determination. Based on frequent monitoring of VO_{2max} and LT, the effectiveness of the training strategies can be analyzed. Furthermore, these measures allow for the setting of individual intensity zones, representing the basis of endurance training.[2] Smartwatches therefore constitute a powerful tool for monitoring, tracking, planning training, races, and recovery. Especially in endurance disciplines like running, cycling, swimming, or triathlon, wearables are used daily. In 2022, Helsen et al reported that in a sample of 2146 active runners, as much as 84.8% used a smartwatch during weekly exercise.[6]

The VO_{2max} estimate of smartwatches from the Garmin® Fenix family (Garmin Deutschland GmbH, Garching) uses extrapolation of running speed and individual HR based on an algorithm developed by Firstbeat Technologies, Finland.[7] To determine the LT, the previously calculated VO_{2max} estimate, running speed acquired by GPS, and HR measured with a chest strap were combined.

If accurate, LT identification using smartwatches can alter the performance testing field. However, the question remains whether these wearables can detect individual thresholds that are valid and reliable enough to replace repeated laboratory-based performance tests, which in Germany, on average, cost around 190–395 € (own market research). Before relying on sports success based on the metrics estimated by a smartwatch, independent research should ensure their accuracy and uncover their potential weaknesses. As the underlying algorithms for wearables are constantly evolving, the need for further evaluation in field tests in step with actual practice is evident.[8] This study aimed to compare the estimates of individual running pace and HR of the Garmin Fenix 7® to individual running pace and HR obtained during a standardized capillary blood lactate field test using the modified Dmax method for analyzation.[9]

In the corresponding literature, previous Garmin Fenix® models have been tested for accuracy in complex physiological measures (Table 1). Wagner et al investigated the reliability of the Garmin Fenix 5® in detecting energy expenditure and VO_{2max} in a sample of 34 male runners.[10] They concluded that the measures were unreliable compared to the corresponding

Table 1 Corresponding Literature on the Accuracy of Physiological and Biomechanical Estimates of Different Garmin® Smartwatch Models

Authors	Smart Watch	Variables Tested	Reference Method	Results	Conclusion
Anderson et al 2019[14]	Garmin® Fenix 5x	VO_2max	Bruce protocol with spiroergometry	$t = 2.21$, $p = 0.037$	Sign. divergent estimate, maybe still an easy and affordable alternative
Wagner et al 2020[10]	Garmin® Fenix 5	VO_2max, energy expenditure (EE), stress level	Spiroergometry, calorimetry, electrocardiography	VO_2max: $CCC_{Lin} = 0.50$; $p = 0.77$ EE: $CCC_{Lin} = 0.43$ $p = 0.52$ Stress level: m-d-Plot; $p = 0.89$	Garmin fenix 5 estimates not sufficiently concordant
Carrier et al 2020[12]	Garmin® Fenix 3 HR	VO_2max, cadence, ground contact time (GCT), vertical oscillation (VO).	Graded exercise testing with spiroergometry, visual 3D	8.05% overestimation of VO_2max (R = 0.917). MAPE: VO: 26.38%; cadence, GCT and VO_2max < 10%.	Valid VO_2max estimate. Cadence and GTC accurate. VO underestimated.
Carrier et al 2021[11]	Garmin® Fenix 6	Speed at LT, HR at LT, speed at OBLA	Graded exercise testing with blood lactate analysis	MAPE: 8.38%, 6.20%, and 10.41%; CCC_{Lin}: 0.85, −0.03, and 0.79	Underestimation of LT speed, poor agreement of OBLA and HR at LT

reference methods, calorimetry, and ergospirometry. Similar results were reported by Anderson et al, who compared the Garmin Fenix 5x® to the ParvoMedics TrueOne 2400® (PMT) in 17 male and 8 female recreational runners. Carrier et al were concerned with the validation of the LT estimate of the Garmin Fenix 6® in comparison to a treadmill-graded exercise test with blood lactate monitoring. Compared with the lab-based test, Fenix 6® underestimated the LT speed and HR at LT; however, it did not differ significantly.[11] Indeed, this study was based on measures of six subjects; therefore, more robust findings are desirable. Another study by Carrier et al reported a mean VO_{2max} overestimation of 8.05% in the Garmin Fenix 3 HR® model compared to measurements obtained from a laboratory-based graded exercise test.[12] A recent systematic review and meta-analysis by the INTERLIVE Consortium concluded that wearables could generally be applied at the population level. However, improvements must be made before their use in sports and clinical purposes is possible.[13]

Thus, the present study aimed to tie in with the corresponding literature and investigate the LT speed and HR estimate of the latest Garmin® Fenix model, the Garmin Fenix 7®. Therefore, we compared the estimated LT speed and HR with values obtained during a field-based graded exercise test. This should provide insights into potential improvements concerning the accuracy of estimated physiological parameters calculated by wearables, thus enabling athletes, coaches, and researchers to make well-informed decisions on how to implement these wearables in practice, racing, and research.

Methods

Transparency and Openness

The study was conducted following the guidelines of the Declaration of Helsinki and was approved by the Ethics Committee of the University of Bundeswehr Munich (M 23–28). Consent was obtained from all the participants before participation in the study. Anonymized data can be obtained upon request.

Participants

We performed an a priori power analysis using G*Power[15] to estimate the required sample size. To compare the smartwatch's estimated lactate threshold and the threshold according to the regular field test, we used a dependent means *t*-test. We adopted an effect size of 0.5, an alpha error probability of 0.05, and a power of 0.8. These input parameters resulted in a priori recommendations for the 34 participants in this study. The inclusion criteria were: (a) all participants were adults and (b) were healthy. Due to incomplete data (48 %), that is, no lactate threshold was ascertained, a resting lactate level ≥2 mmol/l, or dropout, 26 datasets were completed and incorporated into the statistical analysis. The calculated post-hoc power was sufficient (β= 0.99). Only the differences in HR values between the watch and Dmax methods were low (β = 0.42).

Materials

Participants received a Garmin Fénix 7® smartwatch and a HR chest strap (Garmin HRM-Pro) during the study period. Garmin Fénix 7® is a high-end smartwatch product[16] with lactate threshold pace and HR estimation capability. Like many wearables, Garmin uses a light-based photoplethysmography (PPG) sensor to estimate the HR on the wrist. However, the LT measurement only works in combination with the chest strap HRM-Pro.[17] The HRM-Pro sensor measures with a frequency of 2.4 GHz at 8 dBm nominal. It is licensed under us8386009. The instruction was to use the smartwatch and the chest strap in each training session (see 2.4 Study Design). The smartwatch must be connected to an application. As the use of the corresponding Garmin application *Garmin Connect* is prohibited by the institution owing to data security concerns, we used the provider *Fitrockr*.[18] Theoretically, all aggregated data from the watch were stored on German-based servers and exportable However, we did not analyze the synced data because the watch displayed all the required data. For the subsequent field test, the Lactate Scout+ Solo or the Lactate Scout 4 (EKF Diagnostics, Cardiff, UK) were used to measure blood lactate samples from the earlobe. These portable devices need to absorb 0.2 microliter of capillary blood onto a pre-calibrated enzymatic-amperometric sensor. For each measurement, a fresh sensor is connected to the device. They applied a measurement range of 0.5 to 25 mmol/L. The Lactate Scout+ was tested in a study by Bonaventura et al (2015) and four other portable lactate analyzers. Overall, the results showed that all portable analyzers have a negative bias underestimating the values either in low blood lactate concentrations or in high concentrations, depending on the device. For the Lactate Scout+ the measurement error was 3.5 %. However, the authors suggested that portable analyzers are still reliable for determining the LT and added that devices from one manufacturer can be used interchangeably.

Study Design

The study was divided into four parts: (a) introduction to the smartwatch and design, (b) calibration phase, (c) LT test with the watch, and (d) LT field test with blood lactate. First, all participants were introduced to how and when to use the smartwatch. Here, we helped them synchronize their smartwatch with the *Fitrockr* hub and explained the procedure according to the Fenix 7® manual.[16] The watch uses the profile information entered in the initial setting and VO_{2max} calculation to calculate the lactate threshold. Nevertheless, the participants had to run some sessions with the HR chest strap to determine accurate maximum HR and VO_{2max} calculations. The Garmin® manual does not specify the number of training sessions that must be completed to begin the lactate threshold test. Second, to standardize the validation, we instructed participants to run twice per week for 5 km. 1.) At 70% of the individual's maximum HR (70% HR_{max}), and 2.) at 85% HR_{max}. The use of smartwatches for other exercise activities was prohibited. We calculated the maximum heart rate (HR_{max}) using the standard sports science formula (220 minus age). Third, after five weeks, they should execute the lactate threshold test with the watch. Because this test required a maximum load run, one of the authors was presented to provide first aid if needed. After activating the GPS, the participants started the lactate threshold test on the Garmin smartwatch as follows (Figure 1):

1. From the watchface, press START
2. Navigate to the activity *outdoor running*,
3. Hold MENU
4. Select training
5. Select lactate threshold test > start test following the instructions on the screen.

During the test, the watch displayed both current and intended target HR. The task was to run the given intervals at each displayed target HR for a specific time. After completing the test intervals, the participants were notified by the vibration

Figure 1 Procedure on Garmin watch face: getting to the guided lactate threshold test.

signals from the watch. They had to stop the timer and save the activity. Subsequently, the watch allows users to update their HR zones for future training based on the ascertained lactate threshold HR. However, users can always decide whether to accept or reject the latest lactate threshold calculation. In this study, all the participants were instructed to accept the calculated values. Fourth, the lactate threshold values, *lactate threshold pace*, and *lactate threshold HR* determined by the Fenix 7® were compared with a graded test in the field using blood lactate taken from the earlobe ≥ 48 hours later.

Participants were instructed not to fast before the testing (no food <1 hour before testing). A maximum of one cup of coffee was allowed in the morning before testing. We did not control for the female participants' menstrual cycle phase (follicular or luteal). The participants arrived at the test location in groups of two. Resting lactate was taken from the earlobe and analyzed using Lactate Scout+ Solo or Lactate Scout 4 before the first stage started. We chose a starting velocity of 8 km/h for the male participants and 6 km/h for the female participants. To control velocity, one investigator functioned as a pacemaker using an accelerometer on a bike. Each stage was run 2 x 400 m. After completing the stage, the participants stopped at the test location. Subjective exhaustion was measured using the Borg CR10 scale to ensure maximum load.[19,20] HR was read from the watch display, and blood lactate was measured. A standardized break of 30s was used for this purpose. The Borg CR10 scale indicates the rate of perceived exhaustion (RPE) from zero (no exhaustion) to ten (maximum exhaustion).[19,20] At the end of the break, the next stage, which was one km/h faster, commenced. This procedure was repeated until the individual performance limit was reached for each participant. Immediately after completion, the last lactate probe and heart rate (HR) were measured. The modified Dmax method[9,21] was used to analyze the lactate threshold. The method contains:

1. The first increase in lactate (> 0.4 mmol/l compared to the previous stage) served as a rule for determining the *aerobic threshold*
2. The point on the lactate curve at which the aerobic threshold is determined is related to the lactate termination value.
3. Next, we determined the point on the lactate curve with the maximum perpendicular distance to the connection between the two points.
4. Consequently, the *anaerobic threshold* corresponds to the load at the point determined in step 3.

For a visual workflow, see Figure 2. If the resting lactate value of one participant was ≥ 2 mmol/l, they were excluded from the study, as it could be problematic when calculating the anaerobic threshold using this method. After conducting

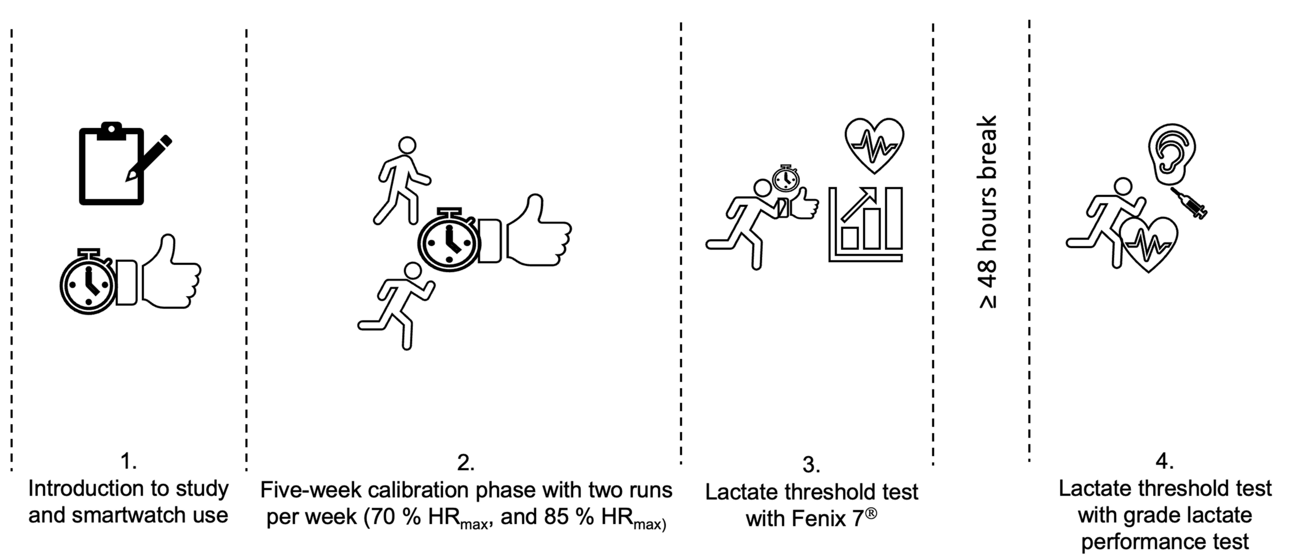

Figure 2 Study Design.

the lactate threshold test once on the watch, the values were updated for each run. Therefore, the field test was tracked as training so that the updated value from the day of the field test, and not from the lactate threshold test of the watch, could be compared with the field test data. Hence, we compared the data obtained from the same day using the watch and the regular used field test with Dmax method.

Statistical Analysis

We used Excel (Microsoft, 2019), SPSS version 29 (IBM Statistics, 2023), and JASP Version 0.16.4 (JASP Team, 2022) for statistical analysis. We compared the ascertained lactate threshold pace of the watch, respectively the km/h, with the pace of the field test (Dmax) as we used this method to calculate the values. In addition, we compared the estimated lactate threshold HR of the watch with the HRs of the field test. All dependent variables, $lactate_{PaceWatch}$, $lactate_{HRWatch}$, $lactate_{PaceDmax}$, and $lactate_{HRDmax}$ were normally distributed, and we used dependent t-tests to determine if the values differed. Additionally, we calculated the mean differences between the two measurements ((Watch-Dmax)/Watch*100) and plotted the mean differences and levels of agreement in a Bland-Altman Plot. All data can be obtained upon request.

Results

Descriptive Statistics

Descriptive data of the 26 participants are listed in Table 2.

Subjective exhaustion was measured using the Borg CR10 scale to ensure maximum load.[19,20] All participants, except for two, reached a maximum exhaustion score of 10. Therefore, we assume all participants reached their individual capacity. We found that the pace at the LT calculated by the Garmin Fenix 7® (M 11.87 km/h ± 1.26 km/h) is 11.96 % lower than the pace at the LT obtained during the field test (M 13.28 km/h ± 1.72 km/h). Similarly, the HR acquired by the Garmin Fenix 7® at LT (174.42 bpm ± 4.62 bpm) is 1.71 % lower than the HR at LT calculated with the Dmax Method (177.42 bpm ± 9.99 bpm). We observed a higher variance in pace and HR at LT for the data gathered during the lactate field test. For an overview, see Tables 3 and 4.

Pace and HR Differences

In general, we observed a significant difference between male and female participants in their maximum km/h at LT as measured by the watch ($t(24)= 4.303$, $p <0.001$, $d = 2.0$) and the Dmax method ($t(24)= 2.802$, $p = 0.01$, $d = 1.3$) (see Figure 3). Garmin underestimated the values for male and female users by 11.48% (male) and 13.58% (female) for the pace at LT, and 1.47% (male) and 2.5% (female) for HR. By comparing the calculated pace at LT of the watch with the value gained during the lactate field test, we observed a significant difference ($t(25)= -6.089$, $p < 0.001$, $d = -1.19$). This equals an average increase of 11.96% in the measured km/h at LT during the lactate field test. Although we also detected

Table 2 Descriptive of Study Sample

	Age	BMI (kg/m²)
Total	25.97 (± 6.26)	24.58 (± 2.8)
Female	27.00 (± 8.58)	23.93 (± 3.06)
Male	25.65 (± 5.57)	24.78 (± 2.75)

Table 3 Descriptive Values of the Dependent Variables as Mean (Std)

	Lactate Pace Watch (km/h)	Lactate Pace Dmax (km/h)	Lactate HR Watch (bpm)	Lactate HR Dmax (bpm)
Total	11.87 (± 1.26)	13.28 (± 1.72)	174.42 (± 4.62)	177.42 (± 9.99)
Female	10.38 (± 1.06)	11.75 (± 1.74)	174.17 (± 5.85)	178.50 (± 11.31)
Male	12.32 (± 0.94)	13.73 (± 1.46)	174.50 (± 4.37)	177.10,178.50 (± 9.861)

Table 4 Differences Between the Measurements for All Participants

Sample	Variable (Watch-Dmax)	Median	Mean (STD)
Total	Difference in Pace (km/h)	−1.56	−1.40 (±1.17)
	Difference Pace (%)	−12.06	−11.96 (± 10.60)
	Difference HR (bpm)	−2.00	−3.00 (±8.33)
	Difference HR (%)	−1.14	−1.71 (±4.78)
Female	Difference in Pace (km/h)	−1.29	−1.37 (±1.64)
	Difference Pace (%)	−11.85	−13.58 (±16.62)
	Difference HR (bpm)	−2.50	−4.33 (±10.21)
	Difference HR (%)	−1.39	−2.50 (±5.90
Male	Difference in Pace (km/h)	−1.56	−1.41 (±1.05)
	Difference Pace (%)	−12.06	−11.48 (±8.62)
	Difference HR (bpm)	−2.00	−2.60 (±7.94)
	Difference HR (%)	−1.14	−1.47 (±4.54)

a 1.71% difference in the HR at LT, the effect was not statistically significant ($t(25) = -1.837$, $p = 0.078$, $d = -0.36$) (see Figure 4). Interestingly, the visualization of all data points and their distribution shows that the differences may be because the smartwatch does not seem to differentiate as accurately in the marginal areas as the Dmax method does for pace and HR. However, smartwatch values accumulated in the middle range (see Figures 5–7).

Discussion

In this study, we aimed to compare the running pace and HR at LT estimated by Garmin Fenix 7® with a standardized graded field test using capillary blood lactate. We found that Fenix 7® underestimated both pace and HR at LT compared to the field test. The mean difference for pace at LT was 11.96%, and 1.71% for the HR. Differences in pace and HR during LT were evident in both genders.

In recent years, wearables have gained popularity because they provide data that can help improve performance or augment healthy habits.[6] Therefore, several studies have investigated the validity and reliability of different variables obtained from wearables, such as VO_{2max} or LT, and compared them to laboratory standards.[13] Similar to our findings,

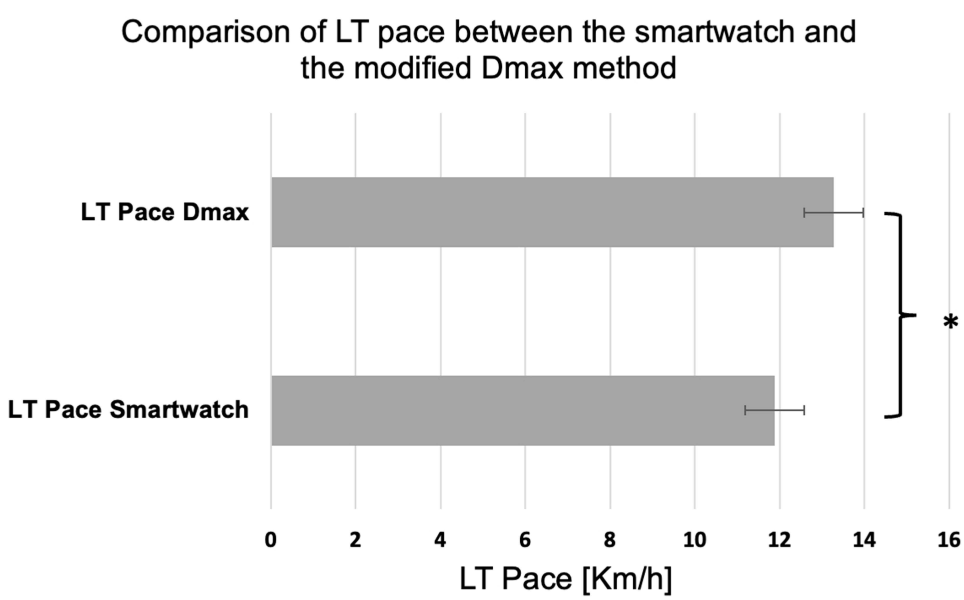

Figure 3 Differences between the LT pace output from the smartwatch and the modified Dmax method.
Notes: Differences between the LT pace output from the smartwatch and the modified Dmax method. Error bars show the SE. * = $p < 0.001$. Graphs were created using Microsoft Excel (Microsoft, 2019).

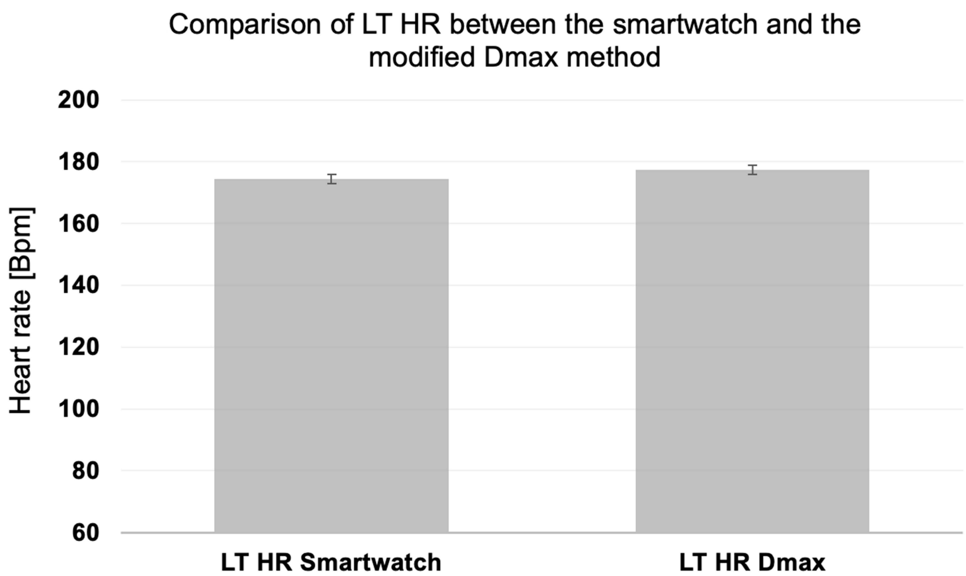

Figure 4 Differences between smartwatch LT HR output (bpm) and the modified Dmax method.
Notes: Differences between smartwatch LT HR output (bpm) and the modified Dmax method. Error bars show the SE. Graphs were created using Microsoft Excel (Microsoft, 2019).

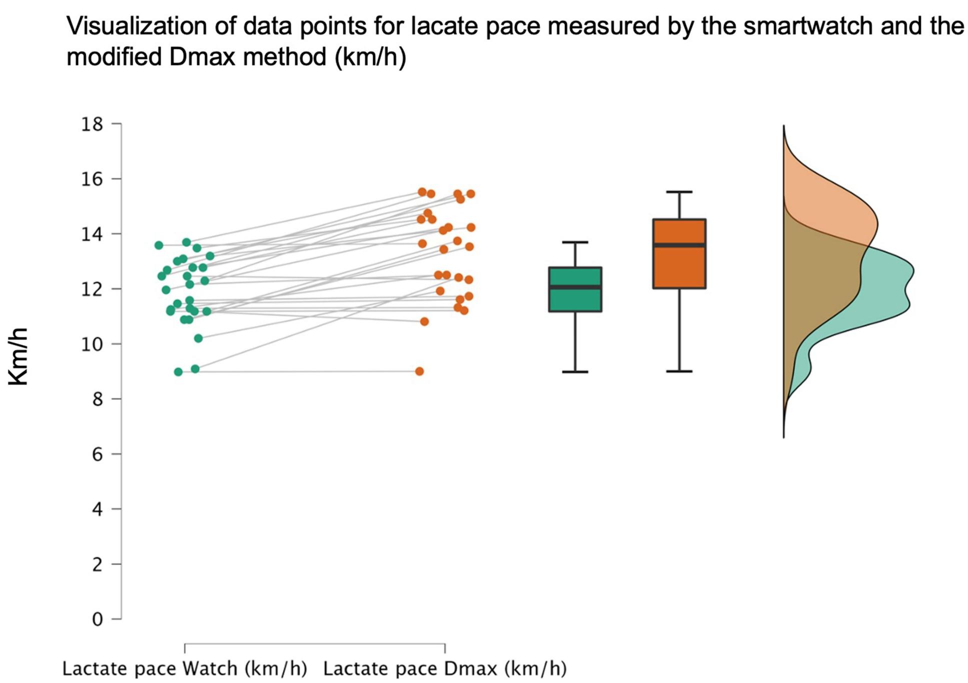

Figure 5 The distribution of all data points (pace).
Notes: The distribution of all data points (26 per comparison group and graph) was illustrated using rain-cloud plots.[22] Here, we compared individual km/h measured with the Garmin Fénix 7® (green) and the Dmax Method (orange), not stratified by sex. The graph was created using JASP (JASP Team, 2022).

they reported deviant values. However, only a few variables can estimate the pace or HR at LT. Hence, existing data are limited. A study by Carrier et al compared Garmin Fenix 6® to a treadmill-based graded exercise test.[11] In line with our results, they found that Garmin Fenix 7® underestimated the pace and HR at LT. However, as the calculated difference was only 8.38% for pace at LT and 6.20% for HR at LT, they concluded that the wearable device was valid. In contrast, we found a greater difference (11.96%) when comparing the estimated pace at LT with that at LT quantified during the field test.

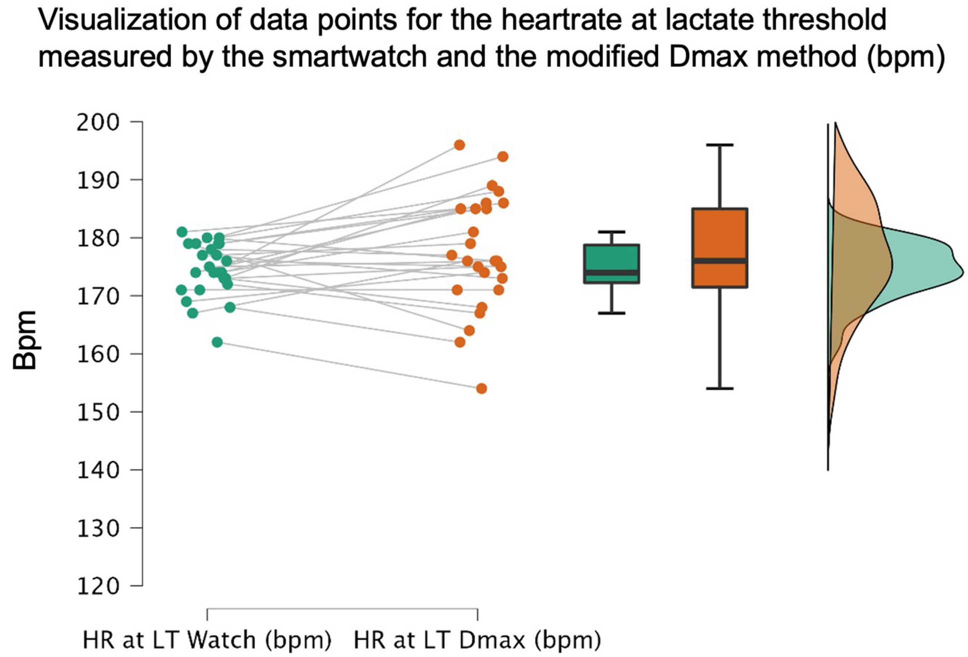

Figure 6 The distribution of all data points (HR).

Notes: All the 26 data points were included in the analysis. Illustrated by rain cloud plots.[22] Here, the comparison of individual HR measured with the Garmin Fénix 7® (green) and the Dmax Method (orange), not stratified by sex. The graph was created using JASP (JASP Team, 2022).

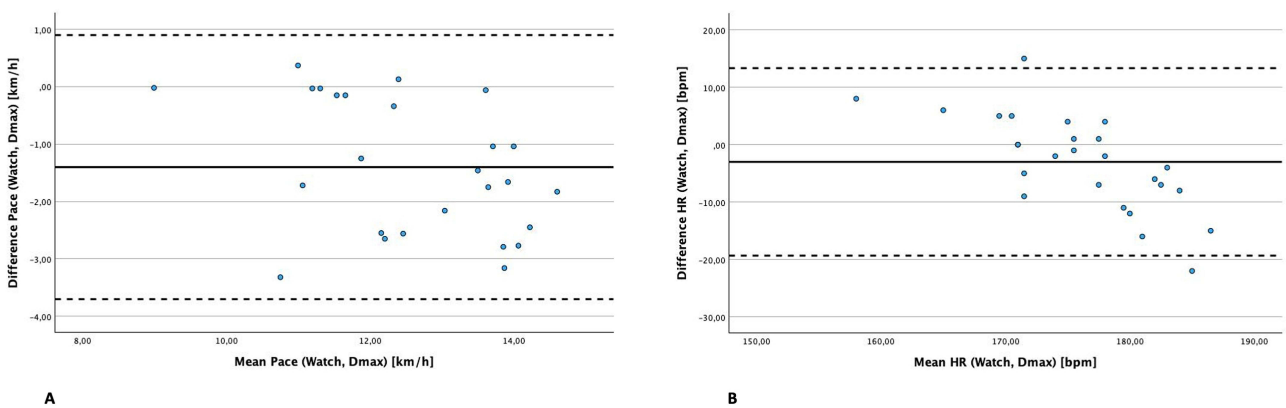

Figure 7 Bland-Altman Plot for differences in pace (**A**) and differences in HR (**B**) plotted with mean values.

Notes: Bland-Altman Plot for differences in pace (**A**) and differences in HR (**B**) plotted with mean values. It shows that the deviations between the measurement methods (here: smartwatch and field test with Dmax method) for pace and HR at the lactate threshold are within the level of agreement.

Interestingly, while the calculated LT values seemed to underestimate the participants' potential, studies on Fenix 5® and 5x® showed the opposite for VO_{2max} estimates. Additionally, Anderson et al (2019) and Wagner et al (2020) reported mean differences of 2.16 mL/kg/min and 4.3 mL/min/kg for the Garmin Fenix 5x® and 5® in recreational runners, respectively.[10,14] These findings are supported by a recent systematic review and meta-analysis that compared the literature on VO_{2max} estimates for 14 different wearables.[13] As there are only some smartwatches that can provide LT estimates, such collective comparisons for this parameter are still missing. However, because the LT estimate is based on a combination of VO_{2max}, pace, and individual HR, it can be expected that the error observed in VO_{2max} estimates will likely be resolved.

For recreational runners, these differences may not cause serious problems. The degree of agreement between LT pace and HR determined with the field test vs estimated by the Garmin Fenix 7® was visualized using a Bland-Altman plot (see Figure 7). For LT pace, most data points are close to the mean without a specific trend. For the HR at LT, however, the two methods seem to deviate with a noticeable trend. At lower HRs, the difference between the means tends to be positive, while at higher HR of >180 bpm, the difference between the means appears to be negative. This underlines the difference in HR variance between the two conditions (see also Figure 6). It is likely that the estimated LT pace and HR still positively influence training, racing, and recovery compared to having no information on these performance metrics. The gathered estimates provide a basis for classifying individual training intensity zones and designing training cycles. If the deviations in LT pace and HR are constant, potential training adaptions, performance gains, or losses can still be detected over time. However, considering a competitive or professional athlete, the same estimation error may have more extensive consequences. The pace at LT was, on average, underestimated by 11.96% by the smartwatch compared to the field test. For example, at an actual LT pace of 4:00 min/km, this deviation can be as high as 28s/km. This distinct underestimation might lead to the permanent application of insufficient or inadequate training stimuli, which would prevent athletes from achieving their best possible performance enhancement within a given training period. In contrast, a potential overestimation of LT pace or HR could lead to unplanned fatigue or over-training symptoms. These assumptions are also supported by the higher variation in the data obtained by the standardized field test. The variance (SD) was greater at both pace and HR at LT when quantified using the field test and modified Dmax method. In this case, a higher variance may indicate a better representation of the inter-individual performance spectra using the modified Dmax method. The lactate field test and subsequent calculation of pace and HR may thus enable scientists and performance coaches to detect finer nuances in performance within and between individuals. This can be used to improve athletic training and consequential performance. Improving performance by paying attention to details and improving minor factors has already been applied by many large teams in competitive sports, such as Team Sky.[23]

Although there is a growing interest in noninvasive measurements to determine LT,[4] data comparing the indirect LT estimates of smartwatches with laboratory tests are scarce. Other indirect concepts use wearable devices with NIRS. In contrast to smartwatches such as Garmin Fenix 6® or 7®, which base the estimate of HR, VO_{2max}, pace, and an algorithm, these devices use optical sensors to quantify the local muscle oxygenation status to obtain the LT. Several studies using NIRS-based principles concluded it to be a low-cost and reliable tool for estimating the lactate threshold, with high correlations for HR and power compared to a capillary blood lactate test.[24–27] Therefore, the assembly of NIRS devices and wearables may represent a promising principle for future developments.[28] As the Garmin® Fenix series can quantify blood oxygenation through optical sensors,[29] a potential way to improve LT estimation could be incorporating blood oxygenation in the calculation.

However, there are some general limitations that impede data interpretation. As previously mentioned, there is a lack of studies testing smartwatch LT estimates in validated laboratory settings or field tests. This is directly linked to a second limitation concerning the study population, which is caused by the lack of available studies and is very homogenous.[24–27] Future studies are needed in the field of competitive sports, recreational sports, and clinical settings across the board. Another concern is the variety of methods applied in studies dealing with wearable devices compared with gold standards. This included both data acquisition and analysis. This problem was addressed by Molina-Garcia et al who presented guidelines for the validation and testing of wearables including reference standards and statistical testing.[13]

Additionally, we want to mention that Garmin® handles their used algorithms completely. This means they can change their calculation bases at any time without providing reasons. This may limit comparability between older and future studies. Another important limitation of the methods applied in this study is the use of the Lactate Scout+ and the Lactate Scout 4 devices for assessing blood lactate during the field test. A study by Bonaventura et al (2015) showed that portable blood lactate analyzers underestimate either low or high values, depending on the device. To the best of our knowledge, today, there is no study on the reliability and accuracy of the Lactate Scout 4 compared to a gold-standard stationary lactate device. Furthermore, this study primarily focused on comparing pace and HR at LT between the two conditions, not the blood lactate values from the field test. Additionally, we did not ask for the female participants' menstrual cycle phase (follicular or luteal).

There is a growing interest in non-invasive methods for determining physiological parameters, such as lactate and the associated thresholds. Various direct and indirect advances have made these variables easily accessible to athletes and coaches. To conclude, supported by our findings and the literature, it seems that the LT estimates calculated by smartwatches, such as Garmin Fenix 7® are not yet comparable to laboratory and standardized field tests. However, in recreational settings, using smartwatches can still serve as a basis for the monitoring and planning of training, racing, and recovery. As the number of available studies in this context is still small, future work is needed to address the current accuracy of LT estimates using smartwatches and ways to improve their underlying principles.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Bundeswehr Munich (protocol code EK UniBw M 23-28 on 18th of May 2023) for studies involving humans.

Informed Consent Statement

Informed consent was obtained from all the subjects involved in the study.

Data Sharing Statement

All anonymized data are available upon request.

Acknowledgments

We would like to thank all subjects for their participation.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This research and the associated Smart Health Lab (SHL) project were funded by the Center for Digitization and Technology Research of the German Bundeswehr. Dtec.bw is funded by the European Union, NextGenerationEU. We acknowledge financial support for open-access publishing by Universität der Bundeswehr München / dtec.bw.

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Alvero-Cruz J, Carnero E, García M, et al. Predictive performance models in long-distance runners: a narrative review. *Int J Environ Res Public Health*. 2020;17(21):8289. doi:10.3390/ijerph17218289
2. Sanders D, Heijboer M, Akubat I, Meijer K, Hesselink MK. Predicting high-power performance in professional cyclists. *Int J Sports Physiol Perform*. 2017;12(3):410–413. doi:10.1123/ijspp.2016-0134
3. Buttar KK, Saboo N, Kacker S. a review: maximal oxygen uptake (VO2max) and its estimation methods. *Internat J Phys Educ*. 2019;2019:24–32.
4. Krishnan A, Guru CS, Sivaraman A, Alwar T, Sharma D, Angrish P. Newer perspectives in lactate threshold estimation for endurance sports - A mini-review. *Central Europ J Sport Sci Med*. 2021;35:99–116. doi:10.18276/cej.2021.3-09
5. Borges NR, Driller MW. Wearable lactate threshold predicting device is valid and reliable in runners. *J Strength Cond Res*. 2016;30(8):2212–2218. doi:10.1519/JSC.0000000000001307
6. Helsen K, Janssen M, Vos S, Scheerder J. Two of a Kind? Similarities and differences between runners and walkers in sociodemographic characteristics, sports related characteristics and wearable usage. *Int J Environ Res Public Health*. 2022;19(15):9284. doi:10.3390/ijerph19159284
7. Firstbeat Technologies Ltd. Automated fitness level (VO2max) estimation with heart rate and speed data; 2017. Available from: https://assets.firstbeat.com/firstbeat/uploads/2017/06/white_paper_VO2max_30.6.2017.pdf. Accessed May 07, 2024.
8. Evenson KR, Spade CL. Review of validity and reliability of Garmin activity trackers. *J Meas Phys Behav*. 2020;3(2):170–185. doi:10.1123/jmpb.2019-0035
9. Cheng B, Kuipers H, Snyder A, Keizer H, Jeukendrup A, Hesselink M. A new approach for the determination of ventilatory and lactate thresholds. *Int J Sports Med*. 1992;13(07):518–522. doi:10.1055/s-2007-1021309
10. Wagner M, Engel F, Klier K, Klughardt S, Wallner F, Wieczorek A. Zur Reliabilität von Wearable Devices am Beispiel einer Premium Multisport-Smartwatch. *German J Exerc Sport Res*. 2021;51(1):49–62. doi:10.1007/s12662-020-00682-7
11. Carrier B, Cruz K, Farmer H, Navalta J. Validation of the lactate threshold estimate from the Garmin fenix 6 fitness tracker. *Med Sci Sports Exerc*. 2021;53(8S):48. doi:10.1249/01.mss.0000759640.87686.98

12. Carrier B, Creer A, Williams LR, et al. Validation of Garmin fenix 3 HR fitness tracker biomechanics and metabolics (VO2max). *J Meas Phys Behav*. 2020;3(4):331–337. doi:10.1123/jmpb.2019-0066
13. Molina-Garcia P, Notbohm HL, Schumann M, et al. Validity of estimating the maximal oxygen consumption by consumer wearables: a systematic review with meta-analysis and expert statement of the INTERLIVE network. *Sports Med*. 2022;52(7):1577–1597. doi:10.1007/s40279-021-01639-y
14. Anderson JC, Chisenall T, Tolbert B, Ruffner J, Whitehead PN, Conners RT. Validating the commercially available Garmin fenix 5x wrist-worn optical sensor for aerobic capacity. *Int J Innov Educ Res*. 2019;7(1):147–158. doi:10.31686/ijier.vol7.iss1.1293
15. Faul F, Erdfelder E, Lang AG, Buchner A, G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39:175–191. doi:10.3758/BF03193146
16. Garmin Ltd. FĒNIX ® 7 SERIES Owner's Manual; 2022. Available from: www.garmin.com. Accessed May 07, 2024.
17. Garmin Ltd. HRM-PRO TM owner's manual; 2019. Available from: www.garmin.com. Accessed May 07, 2024.
18. FitRockr I Health Data research & Analytics platform. Fitrockr health solutions; 2023. Available from: https://www.fitrockr.com/research/. Accessed May 07, 2024.
19. Moura-Fernandes MC, Moreira-Marconi E, de Meirelles AG, et al. Acute effects of whole-body vibration exercise on pain level, functionality, and rating of exertion of elderly obese knee osteoarthritis individuals: a randomized study. *Appl Sci*. 2020;10(17):5870. doi:10.3390/app10175870
20. Shariat A, Cleland JA, Danaee M, et al. Borg CR-10 scale as a new approach to monitoring office exercise training. *Work*. 2018;60(4):549–554. doi:10.3233/WOR-182762
21. Fabre N, Balestreri F, Pellegrini B, Schena F. The modified dmax method is reliable to predict the second ventilatory threshold in elite cross-country skiers. *J Strength Cond Res*. 2010;24(6):1546–1552. doi:10.1519/JSC.0b013e3181dc450a
22. Allen M, Poggiali D, Whitaker K, Marshall TR, van Langen J, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res*. 2021;4:63. doi:10.12688/wellcomeopenres.15191.2
23. Heron N, Usher R, MacLeod D, Sarrieguil I, Mercadel J, Tully MA. Infographics: winning road cycle races: a Team Sky perspective. *Br J Sports Med*. 2018;52(10):633–634. doi:10.1136/bjsports-2017-097819
24. Raleigh C, Donne B, Fleming N. Association between different non-invasively derived thresholds with lactate threshold during graded incremental exercise. *Int J Exercise Sci*. 2018;11:391–403.
25. Farzam P, Starkweather Z, Franceschini MA. Validation of a novel wearable, wireless technology to estimate oxygen levels and lactate threshold power in the exercising muscle. *Physiol Rep*. 2018;6(7):e13664. doi:10.14814/phy2.13664
26. Driller M, Plews D, Borges N. Wearable near infrared sensor for determining an athlete's lactate threshold during exercise. *NIR News*. 2016;27(4):8–10. doi:10.1255/nirn.1609
27. Salas-Montoro JA, Mateo-March M, Sánchez-Muñoz C, Zabala M. Determination of second lactate threshold using near-infrared spectroscopy in elite cyclists. *Int J Sports Med*. 2022;43(08):721–728. doi:10.1055/a-1738-0252
28. Gurel NZ, Jung H, Hersek S, Inan OT. Fusing near-infrared spectroscopy with wearable hemodynamic measurements improves classification of mental stress. *IEEE Sens J*. 2019;19(19):8522–8531. doi:10.1109/JSEN.2018.2872651
29. Lauterbach CJ, Romano PA, Greisler LA, Brindle RA, Ford KR, Kuennen MR. Accuracy and reliability of commercial wrist-worn pulse oximeter during normobaric hypoxia exposure under resting conditions. *Res Q Exerc Sport*. 2021;92(3):549–558. doi:10.1080/02701367.2020.1759768