Temperature

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ktmp20

Reliability of a 60-min treadmill running protocol in the heat: The journal *Temperature* toolbox

Toby Mündel, Simon Gilmour, Marlena Kruger & Jasmine Thomson

To cite this article: Toby Mündel, Simon Gilmour, Marlena Kruger & Jasmine Thomson (2022): Reliability of a 60-min treadmill running protocol in the heat: The journal *Temperature* toolbox, Temperature, DOI: 10.1080/23328940.2022.2143168

To link to this article: https://doi.org/10.1080/23328940.2022.2143168

Published online: 11 Nov 2022.

TEMPERATURE
https://doi.org/10.1080/23328940.2022.2143168

Taylor & Francis
Taylor & Francis Group

METHOD ARTICLE

OPEN ACCESS Check for updates

Reliability of a 60-min treadmill running protocol in the heat: The journal *Temperature* toolbox

Toby Mündel [a], Simon Gilmour [b], Marlena Kruger [c], and Jasmine Thomson [b]

[a]School of Sport Exercise and Nutrition, Massey University, Palmerston North, New Zealand; [b]Fonterra Research and Development Centre, Palmerston North, New Zealand; [c]School of Health Sciences, Massey University, Palmerston North, New Zealand

ABSTRACT

We determined the reliability of a 60-min treadmill protocol in the heat when spaced >4 weeks apart, longer than the test–retest duration of 1 week found in the literature. Nine unacclimated, trained males (age: 31 ± 8 y; VO_2peak: 60 ± 6 ml·kg^{-1}·min^{-1}) undertook a 15 min self-paced time-trial pre-loaded with 45 min of running at 70% of individual ventilatory threshold (11.2 ± 0.3 km·h^{-1}) in 30 ± 1°C (53 ± 5% relative humidity). They repeated this following 40 ± 14 and 76 ± 26 days, with pre-trial standardization of diet and exercise for 48 h. When considering trial 1 as a familiarization, change in core temperature (ΔT_{core}) during the first 45 min (Δ2.0 ± 0.2°C) between trials 2 and 3 yielded bias and 95% limits of agreement (LoA) of −0.10 ± 0.43°C, standard error of measurement (SEM) of 0.13°C and intraclass correlation coefficient (ICC) of 0.75, more reliable than measures of baseline T_{core} (36.9 ± 0.2°C; LoA: −0.23 ± 0.90°C; SEM: 0.22°C; ICC: 0.03) and T_{core} at 45 min during exercise (38.9 ± 0.4°C; LoA: 0.32 ± 1.12°C; SEM: 0.28°C; ICC: 0.15). The coefficient of variation (CV) between trials 2 and 3 for distance run during the 15 min time-trial was 2.1 ± 2.0% with LoA of 0.001 ± 0.253 km and SEM of 0.037 km. This protocol is reliable spaced ~5 weeks apart when considering the most commonly accepted limit of <5% CV for performance, reinforced by reliability of the ΔT_{core} being 0.1 ± 0.4°C.

ARTICLE HISTORY
Received 30 August 2022
Revised 26 September 2022
Accepted 28 September 2022

KEYWORDS
Exercise; heat; performance; core temperature; reliability

Introduction

Performance of, and the core temperature (T_{core}) response to, exercise under heat-stressful conditions are often used as primary outcome measures in studies evaluating treatment effects such as training [1], heat adaptation [2], clothing [3], and diet [4]. Knowledge of the reliability of these measures for a given protocol and sample are vital in understanding whether an outcome measure reflects a "true" treatment effect or simply measurement error. Several studies have utilized a protocol consisting of a time-trial (to measure performance) pre-loaded with a period of fixed-intensity exercise (to measure the T_{core} response) [5–7]. This protocol has been reported as valuable and reliable whether performed in heat-stressful or thermoneutral conditions [5–7], however these studies assessed the reliability of the performance (only) and not the T_{core} response.

Whilst test–retest reliability is usually assessed after a ≤ 7-day period [5–7], and no pre-loaded test has been assessed greater than 2 weeks apart [8], several types of intervention are often administered over a period of 2–4 weeks [1,2] or greater. The greater time between trials increase the chance of factors such as fitness (cf. heat loss effectors) and body composition to change and influence performance and/or T_{core}. Moreover, previous protocols performed under heat-stressful conditions have resulted in (only) moderate heat strain i.e. a rise in T_{core} to ≤39°C [5,6].

Therefore, the purpose of this study was to determine the reliability of a 60 min treadmill protocol under heat-stressful conditions when trials were spaced greater than four weeks apart; specifically, the reliability of the change in T_{core} and time-trial performance.

Materials and methods

This study was approved by the Massey University Human Ethics Committee (Southern A) and conducted in accordance with the latest

CONTACT Toby Mündel t.mundel@massey.ac.nz

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

version of the *Declaration of Helsinki*, with all participants providing informed, written consent. Data were collected in Palmerston North outside of the summer period (i.e. March–November) where the average temperature does not usually exceed 22°C, nor had participants spent any time in a warmer climate for at least one month prior to the study.

Participants

Nine healthy males volunteered to participate, with their mean ± SD physical characteristics being, age: 31 ± 8 y, height: 1.78 ± 0.11 m, weight: 80 ± 13 kg, peak O_2 uptake (VO_2peak): 60 ± 6 ml·kg^{-1}·min^{-1}, maximal heart rate (HR): 185 ± 8 beats·min^{-1} and ventilatory threshold (VT1): 15.9 ± 0.4 km·h^{-1}. Our sample size was guided by previous relevant literature [5,6], and unfortunately influenced/restricted by a period of national lockdowns due to COVID-19 (see *Considerations*). All participants were regularly running more than three times weekly for >20 km·week^{-1} and had participated in running races ≥10 km.

Experimental overview

Participants visited the laboratory on five occasions (Figure 1): 1) preliminary measures and maximal test, 2) experimental familiarization (trial 1), 3) experimental trial (trial 2), 4) re-test of preliminary measures and maximal test, 5) experimental trial (trial 3). Trials 1 and 2 were separated by 40 ± 14 days and trials 2 and 3 were separated by 36 ± 13 days. All trials began at the same time (0900 h) and followed 48 h of dietary and exercise standardization. All trials were completed on an electronically controlled treadmill (True 825, Fitness Technology Inc., MO, USA).

Preliminary measures and maximal test

This visit was conducted in a moderate laboratory environment (20°C, 50% relative humidity). Following weight (Jandever, Taiwan) and height (Seca, Germany) measurements, participants completed an incremental maximal test to determine VO_2peak and VT1 as part of the screening process. The test protocol has been described previously [9]. Briefly, participants commenced the test at 10 km·h^{-1} and 0% gradient. Every minute thereafter the speed increased by 1 km·h^{-1} until the speed reached 18 km·h^{-1}, when the gradient increased by 1% per minute until volitional exhaustion. A fan, located in front of the participants, provided an airflow of 15 km·h^{-1}. HR (Polar, Polar Electro Oy, Finland) and expired gas (TurboFit, VacuMed, Ventura, CA, USA) was

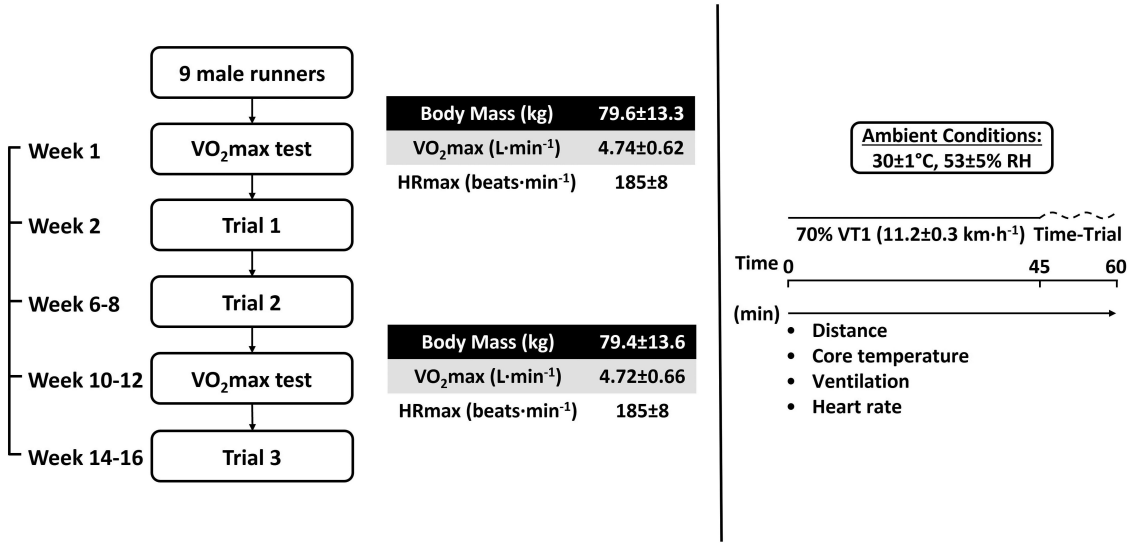

Figure 1. Experimental overview (*left*) and protocol (*right*).

measured continuously. An individual's VT1 was determined by the departure from linearity when plotting ventilation ($L \cdot min^{-1}$) against running speed ($km \cdot h^{-1}$).

Experimental trial development

Two previous investigations using cycle ergometry have successfully utilized a protocol consisting of 45 min of a fixed-intensity pre-load followed by a 15 min time-trial [6,7]. Forty-five minutes of fixed-intensity exercise in the heat is sufficient for measures of T_{core} to change <0.1°C over a 5 min period ("steady-state") [10] and 45–60 min of running is synonymous with 10 km running races. Therefore, we felt this composition to be appropriate. Pilot work identified that 80% of an individual's VT1 [9] did not allow the 45 min duration to be reached, whereas 70% ensured that the full pre-load (45 min) and time-trial (15 min) periods were able to be completed.

Familiarization and experimental trials

These sessions were conducted in a thermally stressful environment of 30 ± 1°C and 53 ± 5% relative humidity, with a fan located in front of the participants with an airflow of 12 $km \cdot h^{-1}$. The treadmill protocol involved 45 min at 70% of individual VT1 followed by a 15 min self-paced time-trial, all at 1% gradient. Participants were provided no feedback except for every 5 min completed, with no external encouragement and only the instruction to complete the time-trial in the fastest time possible. Music was allowed to be played through a loudspeaker using the same personalized playlist and volume for each trial. During the trials, participants were provided with 1.5 $ml \cdot kg^{-1}$ of tap water every 15 min to mimic in-race practice and reduce the risk of participants becoming dehydrated, with whole-body sweat loss estimated by the difference in pre/post body weight accounting for fluid consumed. Measurements taken during the protocol at every 5 min included HR, T_{core} and distance completed whilst expired gas was measured every 15 min for determination of VO_2.

Pre-trial standardization

A 48 h food diary was completed prior to the first trial (familiarization) with participants replicating this for each subsequent trial. Participants were asked to avoid any exercise beyond light/short physical activity (i.e. walking) during the 48 h prior to each trial. Participants were required to attend the laboratory at least 2 h post-prandially having consumed their usual (light) breakfast and at least 500 ml of water. The night prior to the trial participants ingested a radio temperature pill before going to sleep (CorTemp, HQ Inc, Palmetto, FL, USA) as an index of T_{core}. On arrival to the laboratory, participants provided a mid-stream urine sample to confirm urine-specific gravity < 1.020 [11].

Data and statistical analyses

All statistical analyses were performed with SPSS software for Windows (IBM SPSS Statistics 27, NY, USA). Descriptive values were obtained and reported as means and standard deviation (SD) unless stated otherwise. Homogeneity of variance was examined by Levene's test whilst the normality of the data was examined by the Shapiro–Wilk test, with no significant effects. Data analyses were only conducted between the second and third trials due to the known practice effects that occur with repeated testing on account of fitness, skill, or motivation factors that do not usually extend beyond the second trial [8]. Analysis of variance and paired *t*-tests were used to assess differences across trials for distance run and the T_{core} responses, with significance set at $p < 0.05$. Time-trial performance reliability (distance in km) was assessed using several commonly reported measures [12]. Limits of agreement (LoA); [13] are reported as bias±1.96 SD where bias denotes mean difference. The intraclass correlation coefficient (ICC) was calculated based on an absolute agreement, two-way mixed-effects model with demarcations of poor (<0.5), moderate (0.5–0.75, good (0.75–0.90) and excellent (>0.9) reliability [14]. The standard error of measurement (SEM, a.k.a

typical error) was calculated as SD √(1 - ICC) [15]. The within-subject CV was calculated as the SD divided by the mean of two repeated trials then multiplied by 100%. For T_{core} measures, we did not calculate CV as this is contra-indicated due to an arbitrary zero point [16], with studies choosing to report absolute units of measurement (°C) instead.

Our *a priori* acceptance threshold for performance reliability as CV was 5% [17]. An *a priori* acceptance threshold for the T_{core} response was more difficult to rationalize, with few previous studies providing this and/or with similar measures of reliability. Hayden et al. [18] had participants cycle for 60 min at 30% peak work rate in 36°C and reported absolute variations of 0.1°C for exercising rectal temperature based on their measure of CV, whilst Garrett et al. [19] reported absolute variations of 0.15°C for end-exercise rectal temperature based on their measure of CV following cycling for 90 min at 40% peak power in 35°C. Whereas Gant et al. [20] and Ruddock et al [21] provided measures of reliability (bias ± LoA and SEM) but for specific measures of radio pill temperature during a protocol, as opposed to for the protocol *per se*; namely 0.01 ± 0.23°C and 0.08°C [20] and 0.07 ± 0.61°C and 0.12°C [20], respectively. Atkinson and Nevill [12] suggest that a measure is reliable with suitably low bias and narrow 95% LoA. Therefore, we used an *a priori* acceptance threshold for T_{core} responses as 0.1 ± 0.5°C based on the above four studies, due to the precision of the temperature pill used (±0.1°C), and being able to detect differences considered to have physiological consequences or associated with circadian temperature rhythms [22].

Results

Participants began all trials euhydrated based on their urine-specific gravity (1.012 ± 0.007). Participants maintained their body weight and aerobic fitness based on body weight, VO_2peak and maximal HR not being different (all $p > 0.77$) when measured 9–11 weeks apart (Figure 1). Based on their mean VO_2 and HR responses, participants were exercising at an intensity eliciting 72 ± 5% peak VO_2 and 87 ± 5% maximal HR during the 45-min fixed-intensity pre-load. Over the whole exercise period (1 h), participants consumed 0.139 ± 0.146 $L{\cdot}h^{-1}$ water with a whole-body sweat rate of 1.30 ± 0.17 $kg{\cdot}h^{-1}$. None of resting (36.9 ± 0.3°C), end-exercise (39.4 ± 0.5°C) or change in T_{core} (0–45 min: 2.0 ± 0.2°C, 0–60 min: 2.6 ± 0.3°C) were different (all $p > 0.11$) between trials 2 and 3 (Figure 2).

With regard to repeated performance of the time-trials, individual results can be seen in Table 1 and as Bland-Altman plots in Figure 3, with results presenting as homoscedastic. The distance completed for the time-trial during trial 1 was lower than during trial 2 ($p = 0.03$) and trial 3 ($p = 0.04$), which were not different to each other ($p = 0.49$). Correspondingly, the CV decreased from 4.9 ± 3.1% between the first two trials to 2.1 ± 2.0% between the last two trials. Table 2 presents results for mean bias, SEM, and ICC.

Figure 4 displays Bland-Altman plots of T_{core} at baseline, at the end of the fixed-intensity preload (45 min) and the change during the 45 min fixed-intensity preload, with results presenting as homoscedastic. Table 2 presents these results as mean bias, SEM and ICC. Following peer-review, further measures of reliability were added to Table 2 to provide a more complete overview of the current data set for interested readers.

Discussion

This study sought to determine the reliability of the change in T_{core} during a 45 min fixed-intensity

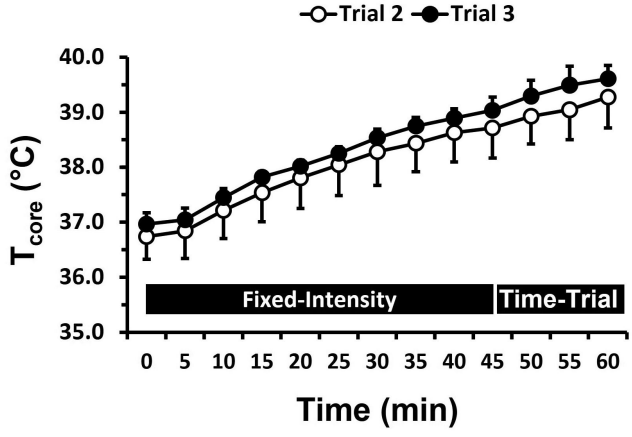

Figure 2. Core temperature (T_{core}) response during the protocol for the second and third trials. Values are mean ± SD.

Table 1. Individual performances of distance completed (km) during the 15-min time-trial.

Participant	Trial 1	Trial 2	Trial 3
1	3.06	3.15	3.26
2	2.82	2.76	2.88
3	2.86	3.18	2.89
4	2.73	3.08	3.10
5	3.17	3.01	2.89
6	2.46	2.80	2.80
7	2.68	2.90	2.93
8	3.15	3.11	3.17
9	2.68	2.87	2.93
Mean ± SD	2.84 ± 0.24	2.98 ± 0.16*	2.98 ± 0.16*

*Significantly different to Trial 1 ($p < 0.05$).

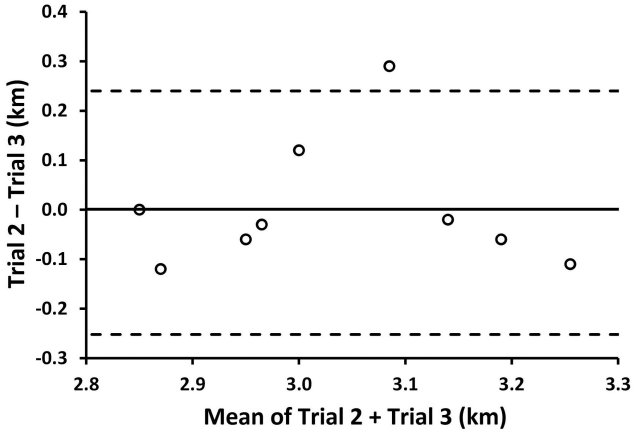

Figure 3. Bland-Altman plot displaying individual time-trial results. Solid black line denotes bias, whilst dashed black lines denote 95% limits of agreement.

pre-load and self-paced performance during the subsequent 15 min time-trial under heat-stressful conditions when trials were spaced approximately 5 weeks apart. The degree of systematic bias (0.10°C and 0.001 km) and random error (95% of results within ±0.43°C and ±0.253 km) are acceptable and demonstrate good reliability of this protocol to measure thermal strain and performance in the same trial.

Measurement error is composed of systematic bias (e.g. learning/practice effect) and random error (e.g. biological variation) that should both be quantified [12]. The most common methods for assessing absolute reliability (CV and SEM) are able to assess systematic bias and are not influenced by the range of values in the sample; unlike the most common methods for assessing relative reliability (*r* and ICC) that should, therefore, not be used to extrapolate results to new individuals or compare between different measurements. Although ICC is widely used in the literature and remains a valuable index in various analysis situations [12,14]. However, whilst absolute measures are expressed dimensionless (CV) or in the actual unit of the measurement (SEM) they represent (only) ~68% of the error present for an individual [12]. Thus, the LoA quantifies systematic bias (0.10°C or 0.001 km) and random error (±0.43°C or ±0.253 km), such that for any new/future participant the difference between their trials should lie within these limits with a ~ 95% probability. This enables researchers and practitioners to correctly assess and interpret the treatment effect of interventions (training, heat adaptation, clothing,

Table 2. Mean ± SD measures between the second and third trials for distance completed during 15-min self-paced time-trials, measures of core temperature (T_{core}), heart rate (HR), rate of O_2 consumption (VO_2) and whole-body sweat rate (WBSR), including mean difference (bias), standard error of measurement (SEM) and intraclass correlation coefficient (ICC). * significant at $p < 0.05$ following analysis of variance.

	Trial 2	Trial 3	Bias ± SD	SEM	ICC
Time-Trial Distance (km)	2.984 ± 0.156	2.983 ± 0.155	0.001 ± 0.129	0.037	0.81*
Baseline T_{core} (°C)	36.96 ± 0.21	36.74 ± 0.42	−0.23 ± 0.46	0.22	0.03
45 min T_{core} (°C)	39.04 ± 0.24	38.71.±0.55	0.32 ± 0.57	0.28	0.15
0–45 min ΔT_{core} (°C)	2.07 ± 0.20	1.98 ± 0.29	−0.10 ± 0.22	0.13	0.75*
0–45 min HR (beats·min^{-1})	168 ± 8	163 ± 12	4 ± 10	3	0.72*
0–45 min VO_2 (L·min^{-1})	3.614 ± 0.319	3.548 ± 0.310	0.066 ± 0.114	0.012	0.96*
0–60 min WBSR (L)	1.308 ± 0.198	1.301 ± 0.158	0.006 ± 0.138	0.038	0.81*

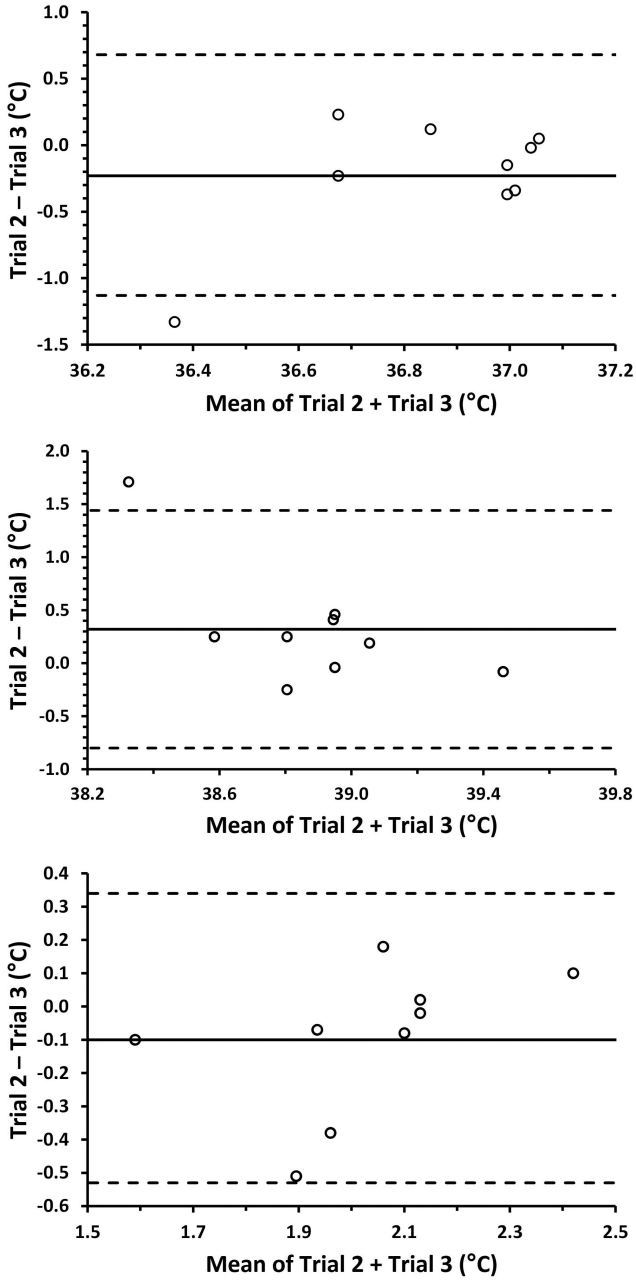

Figure 4. Bland-Altman plots displaying the individual T_{core} at baseline (*upper*), T_{core} at the end (45 min) of the fixed-intensity preload (*middle*) and change in T_{core} during the 45 min fixed-intensity protocol (*lower*). Solid black line denotes bias, whilst dashed black lines denote 95% limits of agreement.

diet, etc.) even when the test-retest is spaced 4–6 weeks apart.

Our protocol induced significant exertional hyperthermia with T_{core} measured at 45 min (38.9 ± 0.4°C, Δ2.0 ± 0.2°C) and end-exercise (39.4 ± 0.5°C, Δ2.6 ± 0.3°C) considerably greater than previous investigations by ≥0.5°C [5,6], although differences in exercise modality are acknowledged (cycling vs. running). Further, by design, the number of data points included in our analyses was low (n = 9). Yet, it is worthwhile noting that compared to previous studies that have collected far greater data volume and/or used post-collection data "cleaning" [20,21], and the suggested physiologically meaningful acceptance threshold of ±0.49°C based on the typical day-to-day variability in resting rectal temperature when controlling for time of day [22], our protocol provides good (or better) reliability.

Our results (CV = 2.1%) for the 15 min time-trial performance compare favorably with the results of Tyler and Sunderland [5] who observed a CV of 2.7% following 75 min treadmill running at 60% VO_2max proceeded by a 15 min self-paced time-trial under similarly warm conditions (30°C, 53% relative humidity), and Che Jusoh et al. [6] who observed a CV of 3.6% following cycle ergometry at 55% VO_2peak for 45 min proceeded by a 15 min time-trial in 29°C and 90% relative humidity. All studies have in common trained and experienced participants that were familiarized with the protocol, rigorous experimental control and standardization of diurnal variation and metabolic/hydration status, minimal participant feedback and a face-valid known end-point to the exercise.

Practical application

The current results have important practical application. Firstly, it should be possible, and highly appropriate/encouraged, to use the known measurement error (for both performance and T_{core}) to determine the validity of intervention success. For example, we have previously demonstrated when using a similar protocol that a dietary intervention resulted in an attenuated T_{core} response of 0.2 ± 0.5°C during the 45 min fixed-intensity pre-load [23]. Based on the current results (Table 2 and Figure 4) we can interpret this finding as not only statistically valid but greater than the measurement error; in other words, we can be confident of the true effect of the intervention based on the signal-to-noise ratio [24]. Secondly, future studies investigating interventions or comparing different populations now have a known measurement error for ΔT_{core} during exercise that can inform sample size justification [16].

Considerations

These observations are valid only for the current sample, protocol, and conditions. Our decision to use only men was guided by the fact that i) even in trained women hormonal differences brought about by menstrual phase and oral contraceptive pill use cause differences in T_{core}, a criterion measure in the current study, of ~0.2°C that interact with exercise intensity and duration [25,26]; ii) data collection was conducted during a period of national lockdowns due to COVID-19 meaning we were unable to control for the above factor(s) in women. Small data sets (e.g. <40) are more sensitive to heterogeneity [12,24]. Therefore, we caution interpretation of the current results without confirmatory research using larger data sets. Finally, the site of T_{core} measurement should be given consideration [10], especially as measurement/agreement can be affected by temperature gradients along the gastrointestinal tract and modifying effects of food and fluid, amongst others [27].

In summary, the current 60 min treadmill running protocol is demonstrated to be reliable when considering the 45 min fixed-intensity change in T_{core} and performance of the subsequent self-paced 15 min time-trial.

Abbreviations

CV	Coefficient of variation
HR	Heart rate
ICC	Intraclass correlation coefficient
LoA	Limits of agreement
SEM	Standard error of measurement
T_{core}	Core temperature
VO_2	Rate of O_2 consumption
VO_2max	Maximal rate of O_2 consumption
VT1	Ventilatory threshold

Acknowledgments

The authors would like to thank Phoebe Jarman and Adam Miller for data collection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Callaghan Innovation [RM22612]; Fonterra Co-Operative Group [RM22104].

ORCID

Toby Mündel http://orcid.org/0000-0002-4214-8543
Simon Gilmour http://orcid.org/0000-0001-9007-3489
Marlena Kruger http://orcid.org/0000-0002-8646-9672
Jasmine Thomson http://orcid.org/0000-0002-2401-9369

References

[1] Souza-Silva AA, Moreira E, de Melo-Marins D, et al. High intensity interval training in the heat enhances exercise-induced lipid peroxidation, but prevents protein oxidation in physically active men. Temperature. 2015;3(1):167–175. DOI:10.1080/23328940.2015.1132101

[2] Malgoyre A, Siracusa J, Tardo-Dino PE, et al. A basal heat stress test to detect military operational readiness after a 14-day operational heat acclimatization period. Temperature. 2020;7(3):277–289. DOI:10.1080/23328940.2020.1742572

[3] Launstein ED, Miller KC, O'Connor P, et al. American football uniforms elicit thermoregulatory failure during a heat tolerance test. Temperature. 2021;8(3):245–253. DOI:10.1080/23328940.2020.1855958

[4] Hess HW, Tarr ML, Baker TB, et al. Ad libitum drinking prevents dehydration during physical work in the heat when adhering to occupational heat stress recommendations. Temperature. 2022;9(3):292–302. DOI:10.1080/23328940.2022.2094160

[5] Tyler C, Sunderland C. The effect of ambient temperature on the reliability of a preloaded treadmill time-trial. Int J Sports Med. 2008;29(10):812–816.

[6] Che Jusoh MR, Morton RH, Stannard SR, et al. A reliable preloaded cycling time trial for use in conditions of significant thermal stress. Scand J Med Sci Sports. 2015;25(Suppl 1):296–301.

[7] Jeukendrup A, Saris WH, Brouns F, et al. A new validated endurance performance test. Med Sci Sports Exerc. 1996;28:266–270.

[8] Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med. 2001;31(3):211–234.

[9] Shing CM, Peake JM, Lim CL, et al. Effects of probiotics supplementation on gastrointestinal permeability, inflammation and exercise performance in the heat. Eur J Appl Physiol. 2014;114(1):93–103.

[10] Mündel T, Carter JM, Wilkinson DM, et al. A comparison of rectal, oesophageal and gastro-intestinal tract temperatures during moderate-intensity cycling in temperate and

hot conditions. Clin Physiol Funct Imaging. 2016;36 (1):11–16.
[11] Sawka MN, Burke LM, Eichner ER, et al. American college of sports medicine position stand. Exercise and fluid replacement. Med Sci Sports Exer. 2007;39:377–390.
[12] Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26 (4):217–238.
[13] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–160.
[14] Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Philadelphia: F.A. Davis; 2020.
[15] Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1–15.
[16] Caldwell AR, Cheuvront SN. Basic statistical considerations for physiology: The journal *Temperature* toolbox. Temperature. 2019;6(3):181–210. DOI:10.1080/23328940.2019.1624131
[17] Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. Sports Med. 2008;38(4):297–316.
[18] Hayden G, Milne HC, Patterson MJ, et al. The reproducibility of closed-pouch sweat collection and thermoregulatory responses to exercise-heat stress. Eur J Appl Physiol. 2004;91(5–6):748–751.
[19] Garrett AT, Rehrer NJ, Patterson MJ, et al. Errors of measurement for blood parameters and physiological and performance measures after the decay of short-term heat acclimation. J Hum Perf Extreme Environ. 2022;17(1):5.
[20] Gant N, Atkinson G, Williams C. The validity and reliability of intestinal temperature during intermittent running. Med Sci Sports Exerc. 2006;38 (11):1926–1931.
[21] Ruddock AD, Tew GA, Purvis AJ. Reliability of intestinal temperature using an ingestible telemetry pill system during exercise in a hot environment. J Strength Cond Res. 2014;28(3):861–869.
[22] Goodman DA, Kenefick RW, Cadarette BS, et al. Influence of sensor ingestion timing on consistency of temperature measures. Med Sci Sports Exerc. 2009;41 (3):597–602.
[23] Che Jusoh MR, Stannard SR, Mündel T. Physiologic and performance effects of sago supplementation before and during cycling in a warm-humid environment. Temperature. 2016;3(2):318–327. DOI:10.1080/23328940.2016.1159772
[24] Zheng H, Badenhorst CE, Lei TH, et al. Measurement error of self-paced exercise performance in athletic women is not affected by ovulatory status or ambient environment. J Appl Physiol. 2021;131(5):1496–1504.
[25] Lei TH, Stannard SR, Perry BG, et al. Influence of menstrual phase and arid vs. humid heat stress on autonomic and behavioural thermoregulation during exercise in trained but unacclimated women. J Physiol. 2017;595:2823–2837.
[26] Lei TH, Cotter JD, Schlader ZJ, et al. On exercise thermoregulation in females: interaction of endogenous and exogenous ovarian hormones. J Physiol. 2019;597:71–88.
[27] Byrne C, Lim CL. The ingestible telemetric body core temperature sensor: a review of validity and exercise applications. Br J Sports Med. 2007;41(3):126–133.