Validity, Reliability and Inertia of Four Different Temperature Capsule Systems

RUNNING TITLE: Comparison of Temperature Capsule Systems

COEN C.W.G. BONGERS[1]
HEIN A.M. DAANEN[2]
CORNELIS P. BOGERD[3]
MARIA T.E. HOPMAN[1]
THIJS M.H. EIJSVOGELS[1,4]

[1]Radboud Institute for Health Sciences, Radboud university medical center, Department of Physiology, Nijmegen, The Netherlands
[2]Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, The Netherlands
[3]TNO, CBRN Protection / Training and Performance Innovations, P.O. Box 45, 2280 AA Rijswijk, The Netherlands
[4]Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom

Total Word Count: 3,336
Abstract Word Count: 250
Total Number of Figures: 3
Total Number of Tables: 3

Disclosure statement funding: This study was supported by a Sportinnovator grant (ZonMw, 2015). The work of T.M.H.E is supported by a European Commission Horizon 2020 grant (Marie Sklodowska-Curie Fellowship 655502).

Author for correspondence:
Coen C.W.G. Bongers, Department of Physiology (392), Radboud university medical center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands, Tel +31 24 36 14468, Fax +31 24 36 16413, E-mail: Coen.Bongers@radboudumc.nl

ABSTRACT

Purpose. Telemetric temperature capsule systems are wireless, relatively non-invasive and easily applicable in field conditions, and have therefore great advantages for monitoring core body temperature. However, the accuracy and responsiveness of available capsule systems have not been compared previously. Therefore, the aim of this study was to examine the validity, reliability and inertia characteristics of four ingestible temperature capsule systems (i.e. CorTemp, e-Celsius, myTemp and VitalSense).

Methods. Ten temperature capsules were examined for each system in a temperature controlled water bath during three trials. The water bath temperature gradually increased from 33°C to 44°C during Trial 1 and 2 to assess the validity and reliability, and from 36°C to 42°C in Trial 3 to assess the inertia characteristics of the temperature capsules.

Results. A systematic difference between capsule and water bath temperature was found for CorTemp (0.077°C±0.040°C), e-Celsius (-0.081°C±0.055°C), myTemp (-0.003°C±0.006°C) and VitalSense (-0.017°C±0.023°C) ($p<0.010$), with the lowest bias for the myTemp system ($p<0.001$). A systematic difference was found between Trial 1 and Trial 2 for CorTemp (0.017°C±0.083°C, $p=0.030$) and e-Celsius (-0.007°C±0.033°C, $p=0.019$), whereas temperature values of myTemp (0.001°C±0.008°C) and VitalSense (0.002°C±0.014°C) did not differ ($p>0.05$). Comparable inertia characteristics were found for CorTemp (25±4 sec), e-Celsius (21±13 sec) and myTemp (19±2 sec), while the VitalSense system responded more slowly (39±6 sec) to changes in water bath temperature ($p<0.001$).

Conclusion. Although differences in temperature and inertia were observed between capsule systems, an excellent validity, test-retest reliability, and inertia was found for each system between 36°C and 44°C after removal of outliers.

Key words: Core body temperature, gastrointestinal temperature, thermoregulation, thermometer

INTRODUCTION

Major sport events are increasingly organized in extreme environmental conditions, making it more important for athletes to perform well in hot and cold ambient conditions and to monitor their core body temperature from a safety perspective (Tc). Exercise-induced increases in metabolic heat production(1, 2) are known to induce a major physiological challenge to the thermoregulatory system(1, 3). A disbalance between heat production and heat loss causes the core body temperature (Tc) to rise, which may lead to the development of exertional hyperthermia (Tc>40°C), heat related illnesses (i.e. heat exhaustion/heat stroke) and/or a reduction of athletic performance(2, 4, 5). Alternatively, exercise in cold environments could lead to rapid heat loss due to conduction (water), convection (wind) and radiation, which may contribute to the development of hypothermia(6). Hence, accurate assessment of an athlete's Tc is important to assess the presence and magnitude of thermoregulatory strain and to select and apply appropriate cooling or heating techniques for preservation of health and exercise performance(7-9).

The gastrointestinal temperature, measured with ingestible temperature capsules, has been established as a valid surrogate marker for Tc(10-12). Temperature capsule systems are wireless, relatively non-invasive and easily applicable in field based conditions. Although the validity of these temperature capsule systems have been examined(11, 13, 14), different study designs were applied and a substantial variation in accuracy was found (i.e. -0.001-0.27°C). Hence, it is essential to determine which capsule system is superior for assessment of Tc in field conditions.

The aim of this study was to examine the validity, reliability and inertia characteristics of four commercially available ingestible telemetric temperature capsule systems (i.e. CorTemp, e-Celsius, myTemp and VitalSense) in well controlled ex-vivo circumstances using a water bath. Data from this study provide insight in which telemetric capsule system has the

most favorable characteristics for Tc assessment, which could enable researchers and trainers to select the best temperature sensor for their scientific study and/or daily practice.

METHODS

Experimental design

Four different ingestible telemetric temperature capsule systems (CorTemp, e-Celsius, myTemp and VitalSense) were tested in a custom made accurately controlled water bath. The primary outcomes were the validity, test-retest reliability and inertia characteristics of the capsule systems. A total of 10 temperature capsules from a single production batch of each telemetry system were tested during three separate trials. The first and second trial consisted of a similar study protocol and was used to assess the validity and test-retest reliability. The third trial adopted a different protocol and was used to examine the inertia characteristics of the temperature capsules. To reduce any bias caused by environmental factors and to ensure that the capsule systems were evaluated in comparable circumstances, a single temperature capsule for each capsule system was used simultaneously in each trial.

Experimental Setup

An overview of the experimental setup is presented in Supplementary Figure 1 (SDC 1, Overview of the experimental setup) . A thermostat-controlled and distilled water-filled bath (3.5 L) was used in which four highly sensitive and calibrated wired temperature probes (1529 Chube E-4 Thermometer Readout Thermistor, Fluke Hart Scientific, Everett, USA) measured temperature up to 0.00035°C exactly. The average value of these wired temperature sensors represented the temperature of the water bath. In addition, a heater (Fluke Hart Scientific 2100 Temperature Controller, Everett, USA) and stirrer (Heidolph Instruments D91126, type RZR1, Schwabach, Germany) system ensured thermal homogeneity of the water bath. A custom made

holder prevented the sensor reaching the bottom of the water bath or coming into contact with another sensor. The external monitors of each of the telemetric capsule systems were placed around the water bath within a distance range of 0.2 m.

Study protocol

Prior to each experiment, the sensors and external monitors were synchronized to ensure that the measurements occurred simultaneously. In the validity and reliability measurements the water bath temperature gradually increased from 33°C to 44°C, exceeding the physiological range between hypothermia (<35°C) and exertional hyperthermia (>40°C). An automated protocol was programmed to induce a stepwise increase in water bath temperature, resulting in twelve temperature plateaus (33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44°C). For each temperature plateau, three conditions had to be achieved before the protocol could proceed: 1) water bath temperature did not vary >0.02°C during fifty consecutive measurements (5 minutes), 2) the average value of the four independent probes did not vary >0.01ºC during two consecutive measurements, and 3) the change in heater power did not exceed 8% during two consecutive measurements. These conditions ensured stability of the water bath temperature and thereby reliable temperature measurements at each point of measurement. The study protocol was performed twice for each temperature capsule (Trial 1/Trial 2), which allowed us to calculate the validity and test-retest reliability. The water bath temperature was measured every 6 seconds.

In the inertia experiment the water bath temperature gradually increased from 36ºC to 42ºC. At every temperature threshold (36, 37, 38, 39, 40, 41 and 42ºC) the water bath temperature was stabilized for five minutes. Then, the water bath temperature increased by 1ºC in a timeframe of five minutes. This timeframe was constructed to mimic the increase in Tc during high intensity exercise in hot ambient conditions, if no heat can be removed from the

body(2). This study protocol allowed us to calculate the time delay of the temperature measured by the temperature capsule compared to the actual temperature of the water bath during the stepwise heating phase. This time delay is defined as the inertia of the temperature capsule.

Telemetric temperature capsule systems

Characteristics of the ingestible telemetric temperature capsule systems are shown in Table 1. All capsule systems used an external wireless recorder to receive the signal from the temperature capsule via a specific radio frequency. The temperature capsules of CorTemp (HQ Inc., Florida, USA), e-Celsius (BodyCap, Caen, France) and VitalSense (Philips Respironics, Bend, Oregon, USA) were delivered in standby modus and had to be activated before use. The myTemp (myTemp, Nijmegen, Netherlands) capsule is automatically activated by the external recorder, which is also the power supply for the temperature capsule. All temperature capsules were activated directly prior to Trial 1. Furthermore, all measurements were performed in accordance with the manual of the individual capsule systems and the highest sample frequency was used throughout the protocol. The external recorders of all capsule systems stored the data, which were exported to a computer for further analysis using the latest version of available software.

Data processing and Statistical Analysis

The average capsule temperature during the final 150 seconds of each temperature plateau was calculated per telemetric system. Due to differences in sample rate, capsule temperature reflected the average of n=25 consecutive measurements for myTemp, n=15 for CorTemp, n=6 for e-Celsius, and n=10 for VitalSense. Average capsule temperature and water bath temperature were compared for each temperature plateau (33-44ºC). Outliers were defined as observations with a difference >1°C between consecutive measurements and were excluded from further analysis. Furthermore, we addressed the number of measurements with a

difference between consecutive data points between 0.2°C and 1.0°C to get more insight into the consistency of the data.

In order to establish the validity, the Bland-Altman method for assessing the agreement between two methods was used(15). In short, the mean difference (=systematic bias) between the temperature capsule and water bath was assessed using a one-sample T-test. The systematic bias and accompanying 95% Limits of Agreement (LOA) were derived from the Bland-Altman plot(15). Furthermore, the Intraclass Correlation Coefficient (ICC) was calculated for the average of all 10 capsules, to determine the inter-measure agreement(16). The Standard Error of Measurement (SEM) was calculated based on the standard deviation (SD) of the difference between temperature capsules and water bath temperature(17). Furthermore, we conducted a Repeated Measures ANOVA to determine whether the accuracy of the capsule systems was different across temperature plateaus (i.e. 33-44ºC). Differences in accuracy across capsule systems were examined using one-way ANOVA. A similar approach was used to determine the test-retest reliability.

Inertia was assessed as the time delay of the telemetric capsule to reach the same temperature as the water bath after a sudden temperature increase. Inertia was determined at 50% (P50) and 90% (P90) of the increase to each temperature plateau, and the time at which the first observation of the capsule and the water bath exceeded the P50 or P90 temperature was taken. Subsequently, the time to reach P50 and P90 of the capsule system was compared with the time of the water bath to reach P50 and P90, and was defined as the time delay (inertia). As the time delay may be influenced by the accuracy and sample frequency of the capsule, we applied two different correction methods: 1) the systematic bias of the telemetric capsule (i.e. sensitivity data) was subtracted from the recorded values, 2) temperature data was interpolated between subsequent samples to determine the exact time at which P50 and P90 were exceeded. Inertia characteristics were presented as: I) raw data, II) corrected for differences in accuracy,

and III) corrected for differences in accuracy and sample frequency. To examine whether there was an inertia difference per temperature plateau across telemetric capsule systems, a two-way repeated measures ANOVA was performed. One-way ANOVA was used to assess the differences in inertia characteristics at P50 and P90 between the four telemetric capsule systems. Furthermore, time constants of the systems response were determined by exposing a single capsule three times to a step change in temperature between two water baths of 7°C (30 – 37°C). Differences in the systems sampling rates did not allow a very precise determination, however by interpolation of the data the time constants can be determined.

All statistical analyses were performed using SPSS Statistics (Version 20), in which the level of significance was set at $p<0.05$. The systematic bias was reported as mean difference ± SD, unless indicated otherwise.

RESULTS

Missing data and outliers. A total of 40 temperature capsules were investigated: 10 sensors per telemetric capsule system. We experienced difficulties with the activation of n=4 VitalSense telemetric capsules, although the provided instructions were carefully followed. Moreover, n=1 of these VitalSense temperature capsules could not be activated at all and 1 temperature capsule stopped measuring after 43ºC during Trial 2, meaning that data of the 44°C temperature plateau of 44°C is not reported for that temperature capsule. As a result, data from 39 temperature capsules was used for our analyses.

In n=6 from n=9 VitalSense temperature capsules, data was randomly missed throughout the protocol (Trial 1 + 2), representing 1.0% of the total data. n=2 CorTemp capsules and n=1 e-Celsius capsule randomly missed 0.1% of the data, whereas no missing data was reported for the myTemp system (Supplementary Table 1, SDC 2, Missing data and outliers). The CorTemp system appeared to be the only system with outliers ($\Delta T_{capsule}$ >1ºC), which was

randomly present in 4.0% of the total data, ranging from a difference of 1°C to 62.1°C. CorTemp also showed error measurements ($0.2^{o}C < \Delta T_{capsule} < 1^{o}C$) in 4.4% of the total data, whereas these error measurements were not present in the other systems. Outliers and error measurements were both found in all CorTemp capsules.

Validity. After exclusion of outliers, mean differences between capsule and water bath temperature for Trial 1 were 0.077±0.040°C (CorTemp), -0.081±0.055°C (e-Celsius), -0.003±0.006°C (myTemp) and -0.017±0.023°C (VitalSense) (Figure 1), which were significantly different from zero (all p-values≤0.01). Additionally, the myTemp system demonstrated the smallest mean difference, followed by VitalSense, CorTemp and e-Celsius ($p_{capsule\ system}<0.001$). The 95% LOA were ±0.079°C (CorTemp), ±0.108°C (e-Celsius), ±0.013°C (myTemp) and ±0.046°C (VitalSense). The SEM was 0.028°C for CorTemp, 0.039°C for e-Celsius, 0.005°C for myTemp and 0.017°C for the VitalSense system. All capsule systems demonstrated an excellent agreement between capsule and water bath temperature based on the significant ICC of 1.00 (all p-values <0.05). The data of Trial 2 revealed similar outcomes with respect to the mean differences, LOA, SEM and ICC (Table 2). A repeated-measures ANOVA indicated that the mean difference between the e-Celsius, myTemp and VitalSense system and water bath temperature did not drift across temperature plateaus ($p<0.05$). In contrast, a significant decrease in mean difference was found across increasing water bath temperatures for the CorTemp system (p=0.002, Figure 2).

Test-retest reliability. Mean difference between Trial 1 and Trial 2 appeared to be significantly different from zero for CorTemp (0.017±0.083°C, LOA= ±0.162°C, p=0.030) and e-Celsius (-0.007±0.033°C, LOA= ±0.064°C p=0.019) (Figure 3). For myTemp (0.0001±0.008°C, LOA= ±0.016°C) and VitalSense (0.002±0.014°C, LOA= ±0.028°C) the mean difference did not differ significantly from zero (both p-values>0.05). Furthermore, the CorTemp system demonstrated the highest mean difference between Trial 1 and Trial 2 (p=0.001), whereas the other systems

had a comparable mean difference between both trials ($p>0.05$). The SEM was 0.058°C for CorTemp, 0.023°C for e-Celsius, 0.006°C for myTemp and 0.010°C for the VitalSense system. An excellent agreement between Trial 1 and Trial 2 was found for all four capsule systems (ICC=1.00, $p<0.05$).

Inertia. Inertia characteristics are summarized in Table 3. The raw data revealed that the CorTemp system had a significant lower time delay to reach p50 (9±5 seconds) and p90 (10±5 seconds) compared to the other capsule systems, whereas the VitalSense system demonstrated the slowest response (p50= 54±12 seconds, p90= 35±3 seconds; $p<0.001$). After correction for the systematic bias of each capsule system, the myTemp system demonstrated the lowest p50 and p90, followed by the CorTemp and e-Celsius system. The p50 and p90 remained the highest for the VitalSense system ($p<0.001$). Additional correction for sample frequency did not alter inertia characteristics (Table 3). Time constants of the systems response were 22 seconds for myTemp, 28 seconds for e-Celsius, 47 seconds for CorTemp and 48 seconds for VitalSense.

DISCUSSION

This is the first study to compare the validity, reliability and inertia characteristics of all commercially available ingestible telemetric temperature capsule systems. Our well controlled ex-vivo water bath study demonstrates that all temperature capsule systems, are valid and reliable to measure (water) temperature, evidenced by their small systematic biases and a low LOA and SEM after removal of outliers (CorTemp). Furthermore, we found that the CorTemp, e-Celsius and myTemp capsule system demonstrated comparable inertia characteristics, whereas the VitalSense system demonstrated a lower responsiveness to changes in water bath temperature. These findings enable researchers and clinicians to select the telemetric capsule system that best suits their goal, which can improve the safety aspect of doing exercise in a hot and cold environment.

An excellent validity and reliability of a temperature measurement technique is characterized by a 1) low systematic bias (<0.1°C), 2) narrow 95% LOA (maximal ±0.4°C), 3) high ICC (>0.80) with the reference temperature, and 4) low SEM(10, 13, 18). We found a significant systematic bias for all four capsule systems, but the validity and reliability of every capsule system complied with reference criteria for an excellent acceptable level of agreement. Nevertheless, we observed a substantial prevelence of outliers in our raw CorTemp data (4.0%), leading to a high LOA (2.3°C) and violation of accuracy criteria (<0.1°C). Data verification and cleaning are, therefore, needed before CorTemp data can be used appropriately. Furthermore, the decreasing systematic bias with increasing temperatures suggests that the CorTemp system is mainly accurate in normothermic and hyperthermic conditions (36-44°C), but less accurate for hypothermic conditions (33-35°C). Although, the CorTemp system did not met the criteria for an excellent validity for hypothermic conditions, the systematic bias (0.1 - 0.2°C) is still physiologically acceptable. e-Celsius, myTemp and VitalSense were more constant and performed well across the whole temperature range. Furthermore, the intraclass correlation coefficient (ICC) and the standard error of measurement (SEM) were used to assess the reliability(17, 18). An ICC of 1.00 was found for all capsule systems, whereas an ICC of >0.80 is typically considerd as acceptable, with higher values respresenting a better reliability(18). The high ICC of the four capsule systems suggests that the error variance between water bath and capsule temperature and between Trial 1 and Trial 2 are negligible compared to the normal variance of the measurement(19). Addionally, the low SEM for all capsule systems is another indication that there is an excellent agreement between water bath and capsule temperature and between Trial 1 and Trial 2. Therefore, all capsule systems are valid and reliable methods to measure temperature after outliers have been removed.

The responsiveness of the temperature capsules was quantified by the inertia characteristics at p50 and p90. We found that the VitalSense system had the slowest response

(38-39 seconds) to acute changes in temperature compared to the other systems (range: 18-26 seconds). Nevertheless, all systems demonstrated an acceptable responsiveness to changes in temperature. A previous study reported a maximal Tc increase of 1°C per 5 minutes if no heat can be removed from the body(2). An inertia of 18 to 39 seconds is, therefore, physiologically irrelevant. Moreover, the underestimation of Tc measured with a temperature capsule in dynamic and/or quick changing situations is marginal and hardly influences final Tc. Furthermore, the order of the results of the time constants matches the results of the p50 and p90 times corrected for sample frequency. The observed time constants are considered appropriate for the physiological signals measured.

Even though the results of our study may be promising, practical considerations must be taken into account. First, the activation of the VitalSense temperature capsules was hard and one of the capsules (10%) could not be activated at all. Anecdotal evidence from our research groups and our collaborators, confirm the infrequent non-activation problem of VitalSense capsules in other studies, whereas similar problems were occasionally experienced for CorTemp capsules. The sample frequency is also an important distinction between the capsule systems, since the sample frequency can be adjusted for CorTemp and myTemp, while it is fixed and relatively low frequent for e-Celsius and VitalSense. Furthermore, 4% of the raw CorTemp data consisted of outliers (>1°C) and another 4.4% of error measurements (0.2-1.0°C). The CorTemp system is therefore less consistent and the use of the raw data with large intervals between measurements might result in inaccurate values. Finally, the present study used capsules from a single production batch from each capsule system, which limited us to assess batch differences within capsule systems.

For human use, other aspects than the investigated accuracy, test-retest reliability and inertia, also play a role. Tc is the result of the local thermal balance affected by tissue properties and local blood flow(20). Studies comparing different measurement location in the digestive

system showed that absolute temperatures and inertia differ between locations(21, 22). Moreover, the esophageal temperature is ~0.2°C lower during moderate intensity exercise compared to both the gastrointestinal and rectal temperature(21). Additionally, the response time of the esophageal temperature is faster than the gastrointestinal temperature, which in turn was faster than the rectal temperature(21). Ideally, the capsule should be located in the gastrointestinal tract and not in the stomach, which can be achieved by timely swallowing the capsule(12, 23).

In conclusion, significant but small differences were observed across telemetric temperature capsule systems. CorTemp demonstrated outliers and error measurements in 4.0% of the recorded data, while this was virtually absent in all other systems. Nevertheless, an excellent validity and test-retest reliability was found for all systems after removal of outliers. The best test-retest reliability was found for the myTemp and VitalSense system, whereas CorTemp and e-Celsius demonstrated a small, but negligible, systematic difference between Trial 1 and Trial 2. Furthermore, the VitalSense system showed the slowest response to increases in water bath temperature, while the other systems had a comparable time delay.

ACKNOWLEDGEMENTS

The authors want to thank Jasmijn Faber for her excellent help during the study. This study was supported by a Sportinnovator grant (ZonMw, 2015). The work of T.M.H.E is supported by a European Commission Horizon 2020 grant (Marie-Sklodowska-Curie Fellowship 655502). For the remaining authors no conflicts of interest were declared. The results of present study do not constitute endorsement by ACSM. Furthermore, the results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. Kenefick RW, Cheuvront SN, Sawka MN. Thermoregulatory function during the marathon. *Sports Med*. 2007;37(4-5):312-5.
2. American College of Sports M, Armstrong LE, Casa DJ et al. American College of Sports Medicine position stand. Exertional heat illness during training and competition. *Med Sci Sports Exerc*. 2007;39(3):556-72.
3. Hargreaves M. Physiological limits to exercise performance in the heat. *J Sci Med Sport*. 2008;11(1):66-71.
4. Gonzalez-Alonso J, Teller C, Andersen SL, Jensen FB, Hyldig T, Nielsen B. Influence of body temperature on the development of fatigue during prolonged exercise in the heat. *J Appl Physiol (1985)*. 1999;86(3):1032-9.
5. Tatterson AJ, Hahn AG, Martin DT, Febbraio MA. Effects of heat stress on physiological responses and exercise performance in elite cyclists. *J Sci Med Sport*. 2000;3(2):186-93.
6. Burtscher M, Kofler P, Gatterer H et al. Effects of Lightweight Outdoor Clothing on the Prevention of Hypothermia During Low-Intensity Exercise in the Cold. *Clin J Sport Med*. 2012;22(6):505-7.
7. Casa DJ, DeMartini JK, Bergeron MF et al. National Athletic Trainers' Association Position Statement: Exertional Heat Illnesses. *J Athl Train*. 2015;50(9):986-1000.
8. Bongers CC, Hopman MT, Eijsvogels TM. Cooling interventions for athletes: An overview of effectiveness, physiological mechanisms, and practical considerations. *Temperature (Austin)*. 2017;4(1):60-78.
9. Bongers CC, Thijssen DH, Veltmeijer MT, Hopman MT, Eijsvogels TM. Precooling and percooling (cooling during exercise) both improve performance in the heat: a meta-analytical review. *Br J Sports Med*. 2015;49(6):377-84.
10. 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):126-33.
11. Challis GG, Kolb JC. Agreement Between an Ingestible Telemetric Sensor System and a Mercury Thermometer Before and After Linear Regression Correction. *Clin J Sport Med*. 2010;20(1):53-7.
12. Gant N, Atkinson G, Williams C. The validity and reliability of intestinal temperature during intermittent running. *Med Sci Sports Exerc*. 2006;38(11):1926-31.
13. Travers GJ, Nichols DS, Farooq A, Racinais S, Periard JD. Validation of an ingestible temperature data logging and telemetry system during exercise in the heat. *Temperature (Austin)*. 2016;3(2):208-19.
14. Bongers C, Hopman MTE, Eijsvogels TMH. Validity and reliability of the myTemp ingestible temperature capsule. *J Sci Med Sport*. 2017.
15. Bland JM, Altman DG. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet*. 1986;1(8476):307-10.
16. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-8.
17. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30(1):1-15.
18. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26(4):217-38.
19. de Vet HCWT MC, Knol LB. *Measurement in Medicine*. 3rd ed. New York: Cambridge University Press; 2011.
20. Taylor NA, Tipton MJ, Kenny GP. Considerations for the measurement of core, skin and mean body temperatures. *J Therm Biol*. 2014;46:72-101.

21. Mundel T, Carter JM, Wilkinson DM, Jones DA. 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):11-6.
22. Teunissen LP, de Haan A, de Koning JJ, Daanen HA. Telemetry pill versus rectal and esophageal temperature during extreme rates of exercise-induced core temperature change. *Physiol Meas*. 2012;33(6):915-24.
23. Bongers CC, Hopman MT, Eijsvogels TM. Using an Ingestible Telemetric Temperature Pill to Assess Gastrointestinal Temperature During Exercise. *J Vis Exp*. 2015;(104).
24. Bongers CCWGH, M.T.E.; Eijsvogels, T.M.H. . Validity and reliability of the myTemp ingestible temperature pill. *Unpublished data*. 2017.

FIGURE LEGENDS

Figure 1. Raw data **(A)** and data after outlier removal **(B)** mean difference between temperature capsule and water bath temperature for the capsule systems. Data were presented as mean difference ± LOA. * indicates a significant systematic bias.

Figure 2. An overview of the mean difference between capsule and water bath temperature for the twelve discrete temperature plateaus. A separate line was plotted for each temperature capsule system. Data were presented as mean difference ± SD, and * represents a drifted response over the temperature plateaus.

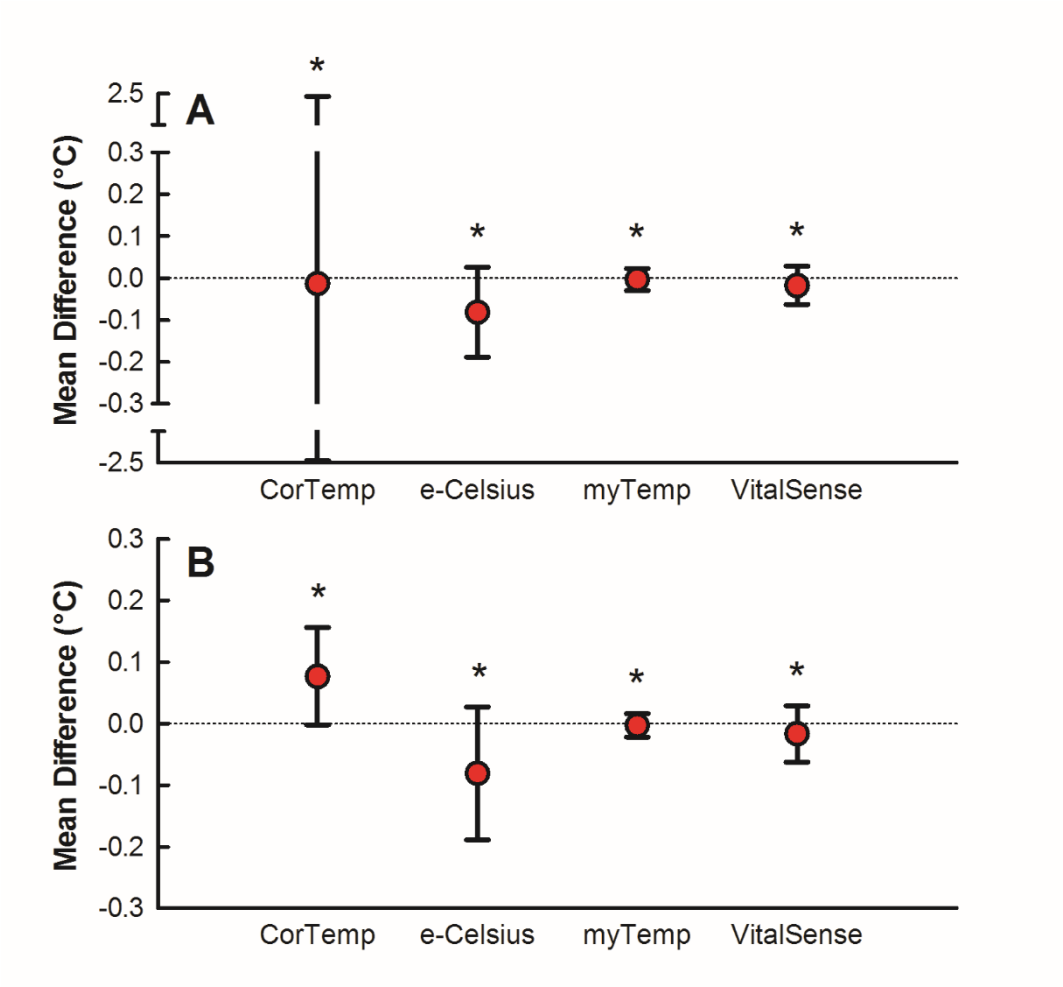
Figure 3. Raw data **(A)** and data after outlier removal **(B)** mean difference between temperatures measured during Trial 1 and Trial 2 for the capsule systems. Data were presented as mean difference ± LOA. * indicates a significant systematic bias.

SUPPLEMENTAL FILES

Supplementary Table 1. Missing data and outliers (Supplementary Table 1.doc)

Supplementary Figure 1. Overview of the experimental setup (Supplementary Figure 1.tiff)

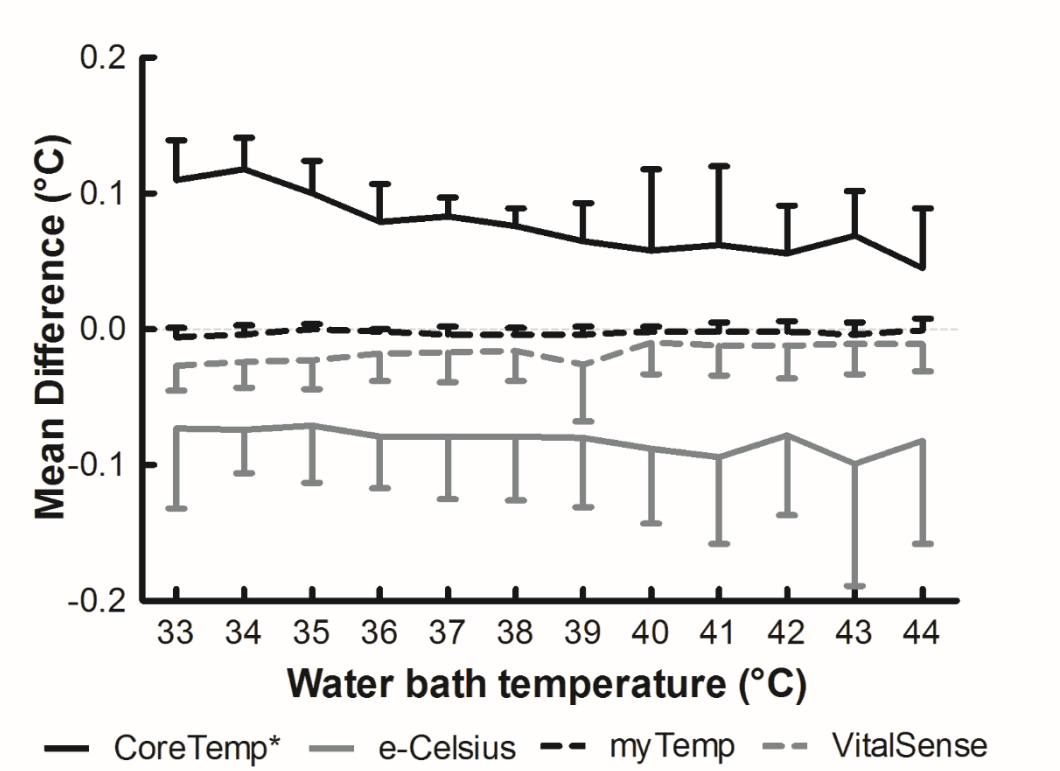
411 **Figure 1.**



412

413

414 **Figure 2.**

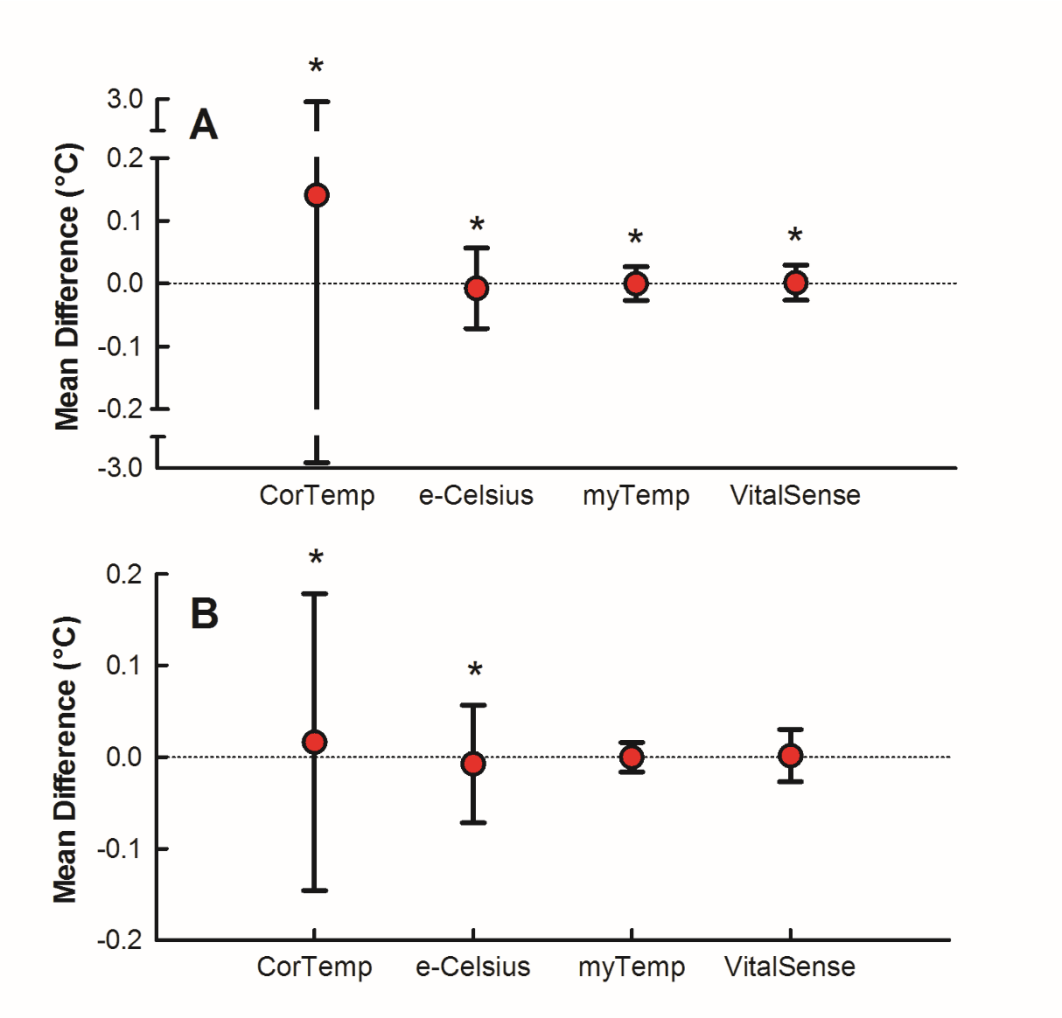


415

416

417

418 **Figure 3.**



419

420

421 **Table 1.**

Table 1. Physical and technical characteristics of the telemetric capsule systems

	CorTemp	e-Celsius	myTemp	VitalSense
Capsule characteristics				
Length (mm)	22.4	17.7	20.0	23.0
Diameter (mm)	10.9	8.9	8.0	8.7
Weight (g)	2.8	1.7	1.3	1.5
Operating range (ºC)	30 to 45	0 to 50	30 to 45	-10 to 60
Accuracy (ºC)	0.27(11)	0.23(13)	0.001(24)	0.17(13)
Battery lifetime	7-10 days	20 days	Infinite	10 days
Power supply	Silver-oxide battery	Zinc-silver oxide battery	Self-induction	Battery
Sample frequency	Adjustable	Fixed	Adjustable	Fixed
Lowest sample rate (sec)	10	~30	6	~15
Software version	CorTrack II	e-Performance manager (v01.01.00.0C)	myTemp Manager (v01.08)	Equivital Manager (v1.2.39.4600)

422

423

424 **Table 2.**

Table 2. Validity of the four temperature capsule systems

		CorTemp	e-Celsius	myTemp	VitalSense
Trial 1	**ICC – raw data**	0.94	1.00	1.00	1.00
	ICC – after outlier removal	1.00	1.00	1.00	1.00
	SEM (°C) – raw data	0.836	0.039	0.005	0.017
	SEM (°C) – after outlier removal	0.028	0.039	0.005	0.017
Trial 2	**MD (°C) – raw data**	-0.154	-0.073	-0.002	-0.018
	MD (°C) – after outlier removal	0.061	-0.073	-0.002	-0.018
	LOA (°C) – raw data	1.466	0.105	0.013	0.037
	LOA (°C) – after outlier removal	0.167	0.105	0.013	0.037
	ICC – raw data	0.98	1.00	1.00	1.00
	ICC – after outlier removal	1.00	1.00	1.00	1.00
	SEM (°C) - raw data	0.529	0.038	0.005	0.013
	SEM (°C) - after outlier removal	0.060	0.038	0.005	0.013

425 **ICC=** Intraclass Correlation Coefficient, **SEM=** Standard Error of the Measurement, **MD=** Mean
426 Difference, **LOA=** Limits of Agreement.

427

428 **Table 3.**

Table 3. Inertia characteristics of the four temperature capsule systems.

		CorTemp	e-Celsius	myTemp	VitalSense	p-value
Raw data	**p50 (s)**	9±5 [b,c,d]	41±17 [a,c]	23±2 [a,b,d]	54±12 [a,c]	**<0.001**
	p90 (s)	10±5 [b,c,d]	27±9 [a,d]	23±3 [a,d]	35±3 [a,b,c]	**<0.001**
Correction I	**p50 (s)**	28±8 [d]	33±12 [c]	22±2 [b,d]	44±7 [a,c]	**<0.001**
(accuracy)	**p90 (s)**	30±6 [d]	33±11 [c,d]	21±1 [b,d]	45±8 [a,b,c]	**<0.001**
Correction II	**p50 (s)**	25±4 [d]	21±13 [d]	19±2 [d]	39±6 [a,b,c]	**<0.001**
(accuracy + sample frequency)	**p90 (s)**	26±7 [d]	21±9 [d]	18±1 [d]	38±9 [a,b,c]	**<0.001**

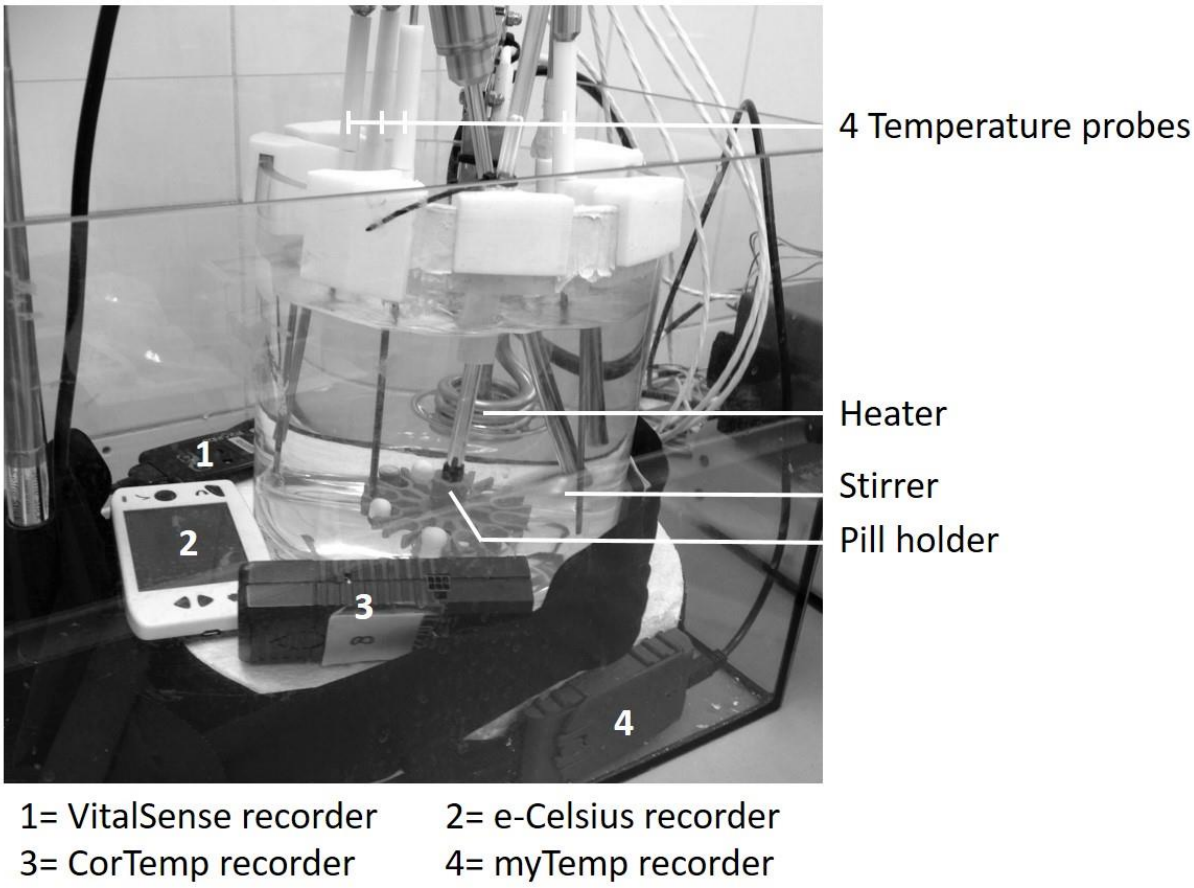
Data were presented as the delay of capsule systems to reach p50 and p90 compared to the water bath. [a] represents significantly different from CorTemp, [b] different from e-Celsius, [c] different from myTemp and [d] different from VitalSense.

429

430

Supplementary Figure 1.

Supplementary Figure 1. Overview of the experimental setup



1= VitalSense recorder 2= e-Celsius recorder
3= CorTemp recorder 4= myTemp recorder

434 **Supplementary Table 1.**

Supplementary Table 1. Missing data and outliers

		CorTemp	e-Celsius	myTemp	VitalSense
Trial 1	**Missing data**	0.1%	0%	0%	0.4%
	Outliers > 1ºC	3.1%	0%	0%	0.1%
	Error measurements $0.2ºC < \Delta T_{capsule} < 1ºC$	4.1%	0%	0%	0%
Trial 2	**Missing data**	0.1%	0.3%	0%	1.5%
	Outliers > 1ºC	4.9%	0%	0%	0.3%
	Error measurements $0.2ºC < \Delta T_{capsule} < 1ºC$	4.7%	0%	0%	0%

435