

Validity and Reliability of Devices That Assess Body Temperature During Indoor Exercise in the Heat

Matthew S. Ganio, MS*; Christopher M. Brown, MA, ATC†; Douglas J. Casa, PhD, ATC, FNATA, FACSM*; Shannon M. Becker, MA, ATC‡; Susan W. Yeargin, PhD, ATC§; Brendon P. McDermott, MS, ATC*; Lindsay M. Boots, BS, ATC*; Paul W. Boyd, BS, ATC*; Lawrence E. Armstrong, PhD, FACSM*; Carl M. Maresh, PhD, FACSM*

*University of Connecticut, Storrs, CT; †Villanova University, Villanova, PA; ‡University of Georgia, Athens, GA; §Indiana State University, Terre Haute, IN

Context: When assessing exercise hyperthermia outdoors, the validity of certain commonly used body temperature measuring devices has been questioned. A controlled laboratory environment is generally less influenced by environmental factors (eg, ambient temperature, solar radiation, wind) than an outdoor setting. The validity of these temperature measuring devices in a controlled environment may be more acceptable.

Objective: To assess the validity and reliability of commonly used temperature devices compared with rectal temperature in individuals exercising in a controlled, high environmental temperature indoor setting and then resting in a cool environment.

Design: Time series study.

Setting: Laboratory environmental chamber (temperature = $36.4 \pm 1.2^{\circ}\text{C}$ [$97.5 \pm 2.16^{\circ}\text{F}$], relative humidity = 52%) and cool laboratory (temperature = approximately 23.3°C [74.0°F], relative humidity = 40%).

Patients or Other Participants: Fifteen males and 10 females.

Intervention(s): Rectal, gastrointestinal, forehead, oral, aural, temporal, and axillary temperatures were measured with commonly used temperature devices. Temperature was measured before and 20 minutes after entering the environmental chamber, every 30 minutes during a 90-minute treadmill walk in the heat, and every 20 minutes during a 60-minute rest in mild conditions. Device validity and reliability were assessed with various statistical measures to compare the measurements using each device with rectal temperature. A device was considered invalid if the mean bias (average difference between

rectal and device temperatures) was more than $\pm 0.27^{\circ}\text{C}$ ($\pm 0.50^{\circ}\text{F}$).

Main Outcome Measure(s): Measured temperature from each device (mean and across time).

Results: The following devices provided invalid estimates of rectal temperature: forehead sticker (0.29°C [0.52°F]), oral temperature using an inexpensive device (-1.13°C [-2.03°F]), temporal temperature measured according to the instruction manual (-0.87°C [-1.56°F]), temporal temperature using a modified technique (-0.63°C [-1.13°F]), oral temperature using an expensive device (-0.86°C [-1.55°F]), aural temperature (-0.67°C [-1.20°F]), axillary temperature using an inexpensive device (-1.25°C [-2.24°F]), and axillary temperature using an expensive device (-0.94°F [-1.70°F]). Measurement of intestinal temperature (mean bias of -0.02°C [-0.03°F]) was the only device considered valid. Devices measured in succession (intestinal, forehead, temporal, and aural) showed acceptable reliability (all had a mean bias = 0.09°C [0.16°F] and $r \geq 0.94$).

Conclusions: Even during laboratory exercise in a controlled environment, devices used to measure forehead, temporal, oral, aural, and axillary body sites did not provide valid estimates of rectal temperature. Only intestinal temperature measurement met the criterion. Therefore, we recommend that rectal or intestinal temperature be used to assess hyperthermia in individuals exercising indoors in the heat.

Key Words: core body temperature, hyperthermia, tympanic membrane

Key Points

- As indicated by mean bias values, invalid estimates of rectal temperature (the “gold standard”) were provided by forehead sticker, oral temperature, temporal temperature, aural temperature, and axillary temperature.
- Intestinal temperature was the only measurement considered valid.
- Successive intestinal, forehead sticker, temporal, and aural measurements demonstrated acceptable reliability.

In a recent study from our laboratory,¹ we assessed the validity of numerous temperature devices to estimate rectal temperature while participants exercised in an outdoor environment. Results indicated that most instruments were invalid. However, the validity of many body temperature devices during indoor exercise is unknown. Core body temperature is often measured during controlled

research experiments when hyperthermia is an outcome variable or a safety criterion. Body temperature also may be measured when athletes participate in indoor sports (eg, basketball, team handball, squash, wrestling, racquetball). Similarly, many athletes use indoor practice facilities during the off-season (eg, strength and conditioning sessions, soccer, football). Accurate measurement of body

temperature indoors is particularly important when exercise is intense or the facility is not air conditioned. Some authorities have speculated that inaccurate measures from temperature devices used during outdoor exercise are due to the influences of cloud cover, wind, rain, and direct sunlight on the device and the person.¹ Compared with the outdoor setting, these same temperature devices may not be as adversely affected in a controlled, consistent indoor environment and, thus, may be more valid. However, they still may not provide valid estimates of core body temperature.

Although rectal temperature (RCT) is the preferred and recommended method of the National Athletic Trainers' Association (NATA) for assessing core body temperature,² athletic trainers use a variety of body sites and devices to measure temperature.³ Common alternatives include oral, aural (tympanic), temporal, and axillary temperature. The validity of these sites in estimating core body temperature has been examined in toddlers,⁴⁻⁶ hospital patients,^{7,8} and participants exercising outdoors.¹ However, many of these authors did not validate measurement sites when hyperthermia was exercise induced. Therefore, it is difficult to extrapolate these findings to healthy individuals exercising indoors.

As recommended by the NATA's position statement on exertional heat illnesses,² RCT is used as the reference standard because of its validity⁹ and practicality in most settings. The evidence indicates that RCT is valid and reliable for individuals at rest and while exercising⁸⁻¹² and is considered the "gold standard" for temperature measurement in hyperthermic athletes.¹³⁻¹⁵ Despite a reported lag in response time (eg, versus esophageal or pulmonary artery temperature), RCT provides a valid and reliable core temperature measurement in the field for diagnosis and treatment of exertional heat stroke.^{13,16-21} It is also an acceptable means of assessing core body temperature in an indoor setting because of its ease of use, low associated cost, and validity.²²

Our purpose was to determine the validity of temperature devices commonly used by sports medicine staff and researchers during indoor exercise in the heat. Indoor exercise was performed in controlled laboratory conditions (ie, an environmental chamber). Based on our previous findings,¹ we expected that devices directly and immediately influenced by changes in skin temperature, the presence of sweat, the ingestion of fluids, or another external influence would not provide the degree of validity and reliability considered acceptable for medical diagnosis or research participant safety monitoring. Specifically, we hypothesized that of the measurements we studied, intestinal temperature, measured via a telemetric pill, would be the only assessment providing a valid estimate of RCT.

METHODS

After signing an informed consent form approved by the institutional review board, which also approved the study, participants completed a self-administered medical history questionnaire and were excluded if contraindications for exercise in the heat were present. All volunteers were physically active (at least 2 workouts or 4 hours of exercise per week). Before testing, we recorded sex, age, height, mass (model BWB-800A; Tanita Corp, Tokyo, Japan), and skinfold thickness at 3 sites to estimate percentage of body fat.²³⁻²⁵

In order to aid in euhydration, participants were instructed to drink 473 mL (16 oz) of fluid both the night before testing and the morning before arrival at the laboratory. We contacted them the night before testing to ensure compliance. Three hours before their scheduled arrival at the laboratory, each volunteer ingested an intestinal thermistor (CorTemp; HQ Inc, Palmetto, FL).

Measurement Sites and Devices

Temperature measurements were taken by a team of 4 trained researchers: one responsible for oral measurements, one for axillary measures, one for aural (ie, ear canal) and temporal measures, and one for recording intestinal (INT), RCT, and forehead sticker (FST) readings. To maintain consistency and accuracy of placement for temporal, aural, and axillary measures, the same researchers measured these sites for each participant.

Rectal temperature was measured with a rectal thermistor (model 401; Yellow Springs Instruments, Yellow Springs, OH) inserted 10 cm past the anal sphincter. Rectal probe calibration was verified by comparing each probe with a certified glass thermometer in water baths of various temperatures (24.5 to 41°C [76.1 to 105.8°F]) and measuring a mean difference of $-0.12 \pm 0.12^{\circ}\text{C}$ ($-0.22 \pm 0.22^{\circ}\text{F}$). The forehead was cleaned with rubbing alcohol and dried, and a sticker (Sportstemp, Greenwood Village, CO) was affixed vertically in the middle of the forehead above the left eyebrow for continuous measurement.

Temporal artery temperature was assessed via a temporal artery scanner (model 2000C; Exergen Corp, Watertown, MA) using the method described in the instruction manual (TEM_{INST}: with no visible sweat, drag along the skin from forehead to hairline; with sweat, hold behind the ear just anterior to the mastoid process) and a modified method observed at local road races (TEM_{MOD}, swipe from forehead to hairline and then around the back edge of the ear ending at just anterior to the mastoid process). Oral temperature was measured using an expensive (SureTemp model 679; WelchAllyn Inc, Skaneateles, NY; ORL_E) and an inexpensive (model VT-801BWT; Walgreens Co, Deerfield, IL; ORL_{IE}) digital thermometer; both were used according to the instruction manuals (tip placed below the tongue, toward the back of the mouth). Aural temperature was measured using a "tympanic" ear thermometer (Thermoscan ExacTemp IRT 4520; Braun, Boston, MA; AUR) according to the instruction manual. Axillary temperature was measured by expensive (DataTherm model 00703; RG Medical Diagnostics, Southfield, MI; AXL_E) and inexpensive temperature devices (model VT-801BWT; Walgreens Co; AXL_{IE}) placed high into the central axillary region, with the volunteer's arm adducted after being wiped free of sweat. Before data analysis, we adjusted AXL_E measures following the procedure described in the instruction manual to estimate rectal temperature: add 1°C [1.8°F]. Thermal sensation was evaluated at each time point using a visual scale (0 = *unbearably cold*, 8 = *unbearably hot*) adapted from Toner et al.²⁶ Participants, while observing the scale, were asked, "How hot or cold do you feel right now?"

In an attempt to examine if cost differences among devices were responsible for various validity and reliability values, an "inexpensive" and an "expensive" model were used for

oral and axillary temperature measurements. Because of patent laws and trade secrets, we do not know the specific technologic differences between an inexpensive and an expensive model. However, we hypothesized that the more expensive device had technologic advances that made it more valid and reliable than its inexpensive counterpart.

Measurements at each site were started at the same time by each researcher. For each researcher, the order of measurement was as follows: Researcher A: RCT, INT, FST; Researcher B: ORL_E, ORL_{IE}; Researcher C: AUR (twice), TEM_{INST} (twice), TEM_{MOD} (twice); Researcher D: AXL_E in left axilla and AXL_{IE} in right axilla.

Measurements took less than 2 minutes except for those with the inexpensive thermometers (ORL_{IE}, AXL_{IE}). These devices took up to 5 minutes to stabilize and provide readings. At the end of all measures, researcher A repeated his or her measures to assess any change in temperature by these devices, and the continuous reading of AXL_E was recorded.

Protocol

Upon arrival, participants inserted a rectal thermistor 10 cm past the anal sphincter, and we applied a forehead sticker to each person. Volunteers were evaluated with 9 different temperature devices after 20 minutes of standing in a cool environment (minute -20; temperature = approximately 23.3°C [74.0°F], relative humidity = 40%) and after 20 minutes of standing in an environmental chamber (minute 0; temperature = 36.4 ± 1.2°C [97.5 ± 2.16°F], relative humidity = 52%). Participants then walked on a treadmill (model TR-9100; Life Stride, Franklin Park, IL) at 5.8 to 6.8 km/h (3.6 to 4.2 mph) at a 5% grade with no fan directed on them. Every 30 minutes, during 90 minutes of exercise, temperature measures were taken during a 5-minute standing break (minutes 30, 60, and 90). Just before temperature measurements, volunteers were moved off the treadmill and away from direct air movement from the fan. After 90 minutes of exercise, they exited the chamber and stood or sat in a cool environment (temperature = approximately 23.3°C [74.0°F]) for an additional 60 minutes; measurements were repeated every 20 minutes (minutes 110, 130, and 150). Participants were allowed to drink water ad libitum throughout the testing protocol except in the 5 minutes before temperature measurement, when no fluid ingestion was allowed.

Statistical Analyses

Temperature Device Validity. Values used for RCT were an average of the RCTs at the beginning and end of the 5-minute temperature-measuring time period. Measures from other devices that were taken twice in a given time period (INT, FST, TEM_{INST}, TEM_{MOD}, AUR) were also averaged when comparing values with RCT.

A 2-way (temperature device × time) repeated-measures analysis of variance was conducted to test the significance of mean differences in devices over time. To evaluate differences in a given device versus RCT, follow-up repeated-measures *t* tests with the Bonferroni α correction were used. Greenhouse-Geisser corrections were made when the assumption of sphericity was violated.

Validity of each device versus RCT was evaluated with a range of measurement error statistics.^{27,28} Mean bias and

limits of agreement were calculated as described by Bland-Altman.²⁹ Briefly, limits of agreement were calculated by multiplying the SD of the mean difference between temperature device and RCT by 1.96 (2 SDs).²⁹ The difference between the measurements with the temperature device and RCT, with a 95% probability, is expected to lie within the limits of agreement.²⁷ Intraclass correlation coefficients, standard error of the mean, and coefficient of variation were calculated as outlined by Atkinson and Nevill.²⁷ Pearson correlation coefficients (*r*), corrected for repeated measures,³⁰ were calculated to evaluate relative agreement of devices. Although limits of agreement, intraclass correlation coefficient, standard error of the mean, coefficient of variation, and *r* provide insight regarding the validity of a device, we determined that, for practicality and accuracy when used by an athletic trainer to assess degree of hyperthermia, mean bias of a given device greater than or less than 0.27°C (0.50°F) from RCT would bring a device's validity into question.¹

Temperature Device Reliability. Device measures that were taken twice during a given time period (INT, FST, TEM_{INST}, TEM_{MOD}, AUR) were evaluated for intradevice reliability. The 1st and 2nd measurements of the same device at a given time point were compared. We calculated measurement error statistics similar to those used to assess temperature validity (ie, mean difference, limits of agreement, intraclass correlation coefficient, standard error of the mean, coefficient of variation, and *r*). All statistical tests were performed with SPSS (version 10 for Windows; SPSS Inc, Chicago, IL) with α set at .05.

RESULTS

Fifteen males and 10 females (mean age = 26.5 ± 5.3 years, height = 174.3 ± 11.1 cm, mass = 72.73 ± 15.95 kg, body composition = 16.2 ± 5.5% body fat) participated in this study. Between 60 and 90 minutes of exercise, 3 volunteers discontinued exercise due to one or a combination of the following termination criteria: RCT greater than 40.00°C, voluntary exhaustion, and observed or self-reported central nervous system dysfunction. Therefore, data at minute 90 are provided for 22 participants.

The interaction of time and temperature device was significant ($F_{7,63} = 24.80$, $P < .001$; Figure 1). From the beginning to the end of each 5-minute temperature measuring period, RCT did not change (averages = 37.94 ± 0.73°C and 37.93 ± 0.73°C, respectively; $F_{1,193} = 0.21$, $P = .649$).

Temperature at Rest. In ambient conditions, differences among devices were evident before exercise began ($F_{9,216} = 30.48$, $P < .001$). The INT, FST ORL_E, ORL_{IE}, AUR, AXL_E, and AXL_{IE} were all different from RCT ($P < .001$). The TEM_{INST} and TEM_{MOD} were the only devices providing measurements that did not differ from RCT ($P = .148$ and $.181$, respectively). After 20 minutes in the heat (minute 0), FST ($P < .001$), TEM_{INST} ($P = .006$), TEM_{MOD} ($P < .001$), ORL_{IE} ($P < .001$), AXL_E ($P = .004$), AUR ($P = .040$), and AXL_{IE} ($P < .001$) were different from RCT. The INT and ORL_E did not differ from RCT ($P = .414$ and 1.000 , respectively).

Temperature During Exercise. At each exercise time point (minutes 30, 60, and 90), TEM_{INST}, TEM_{MOD}, ORL_E, ORL_{IE}, AXL_E, AXL_{IE}, and AUR were all lower than RCT ($P < .001$ -.011). The FST was greater than RCT

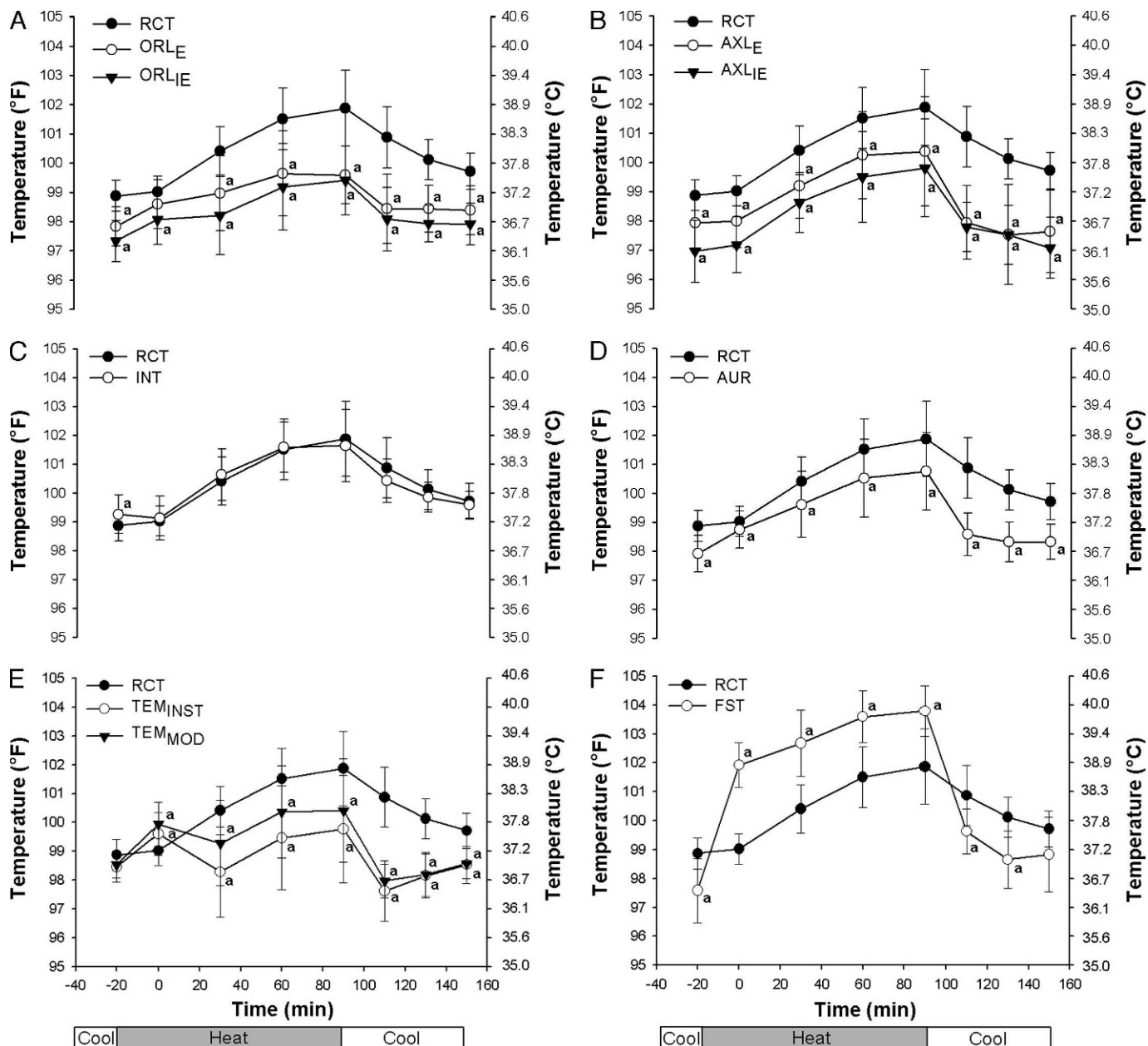


Figure 1. Mean \pm SD of each temperature device over time compared with rectal temperature (RCT). ORL_E indicates oral temperature with expensive thermometer; ORL_{IE}, oral temperature with inexpensive thermometer; AXL_E, axillary temperature with expensive thermometer; AXL_{IE}, axillary temperature with inexpensive thermometer; INT, intestinal temperature; AUR, aural temperature; TEM_{INST}, temporal temperature measured with the method described by the instructional manual; TEM_{MOD}, temporal temperature measured in a modified method; FST, forehead sticker temperature. (See text for further descriptions.) ^a Indicates difference from RCT at the same time point ($P < .05$).

at all exercise time points ($P < .001$). The INT and RCT were not different at 30, 60, and 90 minutes ($P = 1.00$).

Temperature Postexercise. At each postexercise time point (minutes 110, 130, and 150), TEM_{INST}, TEM_{MOD}, ORL_E, ORL_{IE}, AUR, AXL_E, and AXL_{IE} were all lower than RCT ($P < .001$). The FST was less than RCT at minutes 110 and 130 ($P < .001$) but not at minute 150 ($P = .543$). The INT and RCT were not different at 110, 130, and 150 min ($P = .607, 1.00$, and 1.00 , respectively).

Temperature Device Validity. Mean bias, r , limits of agreement, intraclass correlation coefficient, standard error of the mean, and coefficient of variation are presented in Table 1. Mean bias and limits of agreement are represented graphically with Bland-Altman plots in Figure 2.

Temperature Device Reliability. Mean difference between measures, r , limits of agreement, intraclass correlation coefficient, standard error of the mean, and coefficient of variation for INT, FST, TEM_{INST}, TEM_{MOD}, and AUR are presented in Table 2.

DISCUSSION

Our purpose was to evaluate the reliability and validity of selected body temperature-measuring devices commonly used to estimate internal body temperature. The recommended method for the evaluation of internal body temperature in the diagnosis of exertional heat stroke in a field setting is RCT.^{2,15} Previous research¹ from our

Table 1. Measures of Validity Using Rectal Temperature as the Criterion Standard

Temperature Measurement	Bias	Pearson Correlation Coefficient, <i>r</i>	Limits of Agreement	Intraclass Correlation Coefficient	Standard Error of the Mean	Coefficient of Variation, (%) ^a
Intestinal temperature	−0.02°C −0.03°F	0.86	±0.82°C ±1.48°F	0.90	0.22°C 0.40°F	1.84% (C) 1.25% (F)
Forehead temperature	0.29°C 0.52°F	0.58	±2.27°C ±4.08°F	0.62	0.68°C 1.23°F	2.91% (C) 1.99% (F)
Temporal measurement						
Per instruction manual	−0.87°C −1.56°F	0.25	±1.77°C ±3.18°F	0.44	0.65°C 1.17°F	2.32% (C) 1.57% (F)
Modified method	−0.63°C −1.13°F	0.38	±1.65°C ±2.96°F	0.56	0.54°C 0.98°F	2.18% (C) 1.48% (F)
Oral temperature						
Inexpensive device	−1.13°C −2.03°F	0.70	±1.09°C ±1.96°F	0.81	0.39°C 0.70°F	2.40% (C) 1.63% (F)
Expensive device	−0.86°C −1.55°F	0.60	±1.24°C ±2.24°F	0.73	0.42°C 0.76°F	2.17% (C) 1.47% (F)
Aural temperature	−0.67°C −1.20°F	0.77	±1.02°C ±1.84°F	0.86	0.30°C 0.55°F	2.16% (C) 1.46% (F)
Axillary temperature						
Inexpensive device	−1.25°C −2.24°F	0.77	±1.10°C ±1.98°F	0.86	0.38°C 0.68°F	2.71% (C) 1.84% (F)
Expensive device	−0.94°C −1.70°F	0.60	±1.59°C ±2.87°F	0.70	0.53°C 0.96°F	2.59% (C) 1.75% (F)

^a Value (C) = coefficient of variation when using °C; value (F) = coefficient of variation when using °F. See text for further descriptions.

laboratory indicated that many commonly used temperature devices are invalid in an outdoor setting. We examined if controlling environmental factors (eg, sun, wind) would decrease the variability and, hence, increase the performance of these devices. Our stated hypotheses were supported in that, when compared with RCT, only the ingestible temperature device provided a viable means of measuring internal body temperature in individuals who become hyperthermic (average = 38.80 ± 0.72°C [101.84 ± 1.30°F]) during exercise in a controlled, indoor environment.

Oral Temperature

At rest, in mild air conditions (minute −20) ORL_E and ORL_{IE} were both lower than RCT. The mean differences between ORL_E and ORL_{IE} and RCT at rest were −0.58°C and −0.85°C, respectively. Overestimation of core temperature by oral device has been reported in the range of 0.12°C to 0.53°C.^{8,31} After the participant stood for 20 minutes in a hot environment (36.6°C [97.9°F]), RCT did not increase (*P* = .096), but ORL_E and ORL_{IE} increased 0.78°C (*P* = .006) and 0.75°C (*P* = .019), respectively. These changes decreased the differences between ORL_E or ORL_{IE} and RCT (Figure 1). Because RCT did not change when the participant moved to a warmer environment and ORL_E and ORL_{IE} did, our data supported the hypothesis that, in this setting, oral temperature was influenced by ambient temperature.

During and after exercise, when RCT was greater than resting values, ORL_E and ORL_{IE} were lower than RCT at every time point measured. Of the devices used in this study, at peak RCT (minute 90: 38.80 ± 0.72°C [101.84 ± 1.30°F]), ORL_E and ORL_{IE} had the largest mean differences from RCT (−1.25 ± 0.72°C [−2.25 ± 1.30°F] and −1.35 ± 0.45°C [−2.43 ± 0.81°F], respectively). Others^{31,32} have reported differences between oral temper-

ature and RCT during exercise of −0.55°C to −0.33°C. Differences may be attributed to the low peak RCT reached in these studies (37.57°C–37.64°C).^{31,32} Mairiaux et al³¹ reported that the difference between RCT and oral temperatures increased as RCT increased.

Compared with oral temperature measurement during outdoor exercise,¹ ORL_E and ORL_{IE} measured indoors more closely agreed with RCT. For example, Casa et al¹ reported mean differences of −1.20°C (−2.17°F) and −1.67°C (−3.00°F), between the same 2 devices used in this study (ORL_E and ORL_{IE}) and RCT. Despite the closer agreement in the present study (Table 1), ORL_E and ORL_{IE} were still considered invalid for estimated RCT, using the cutoff ± 0.27°C (0.50°F). A change in mean difference between these devices and RCT in a variety of settings further supports the hypothesis that oral temperature is influenced by environmental factors and is not a valid measurement for estimating core body temperature. Also in the present study, ORL_E measured slightly higher than ORL_{IE} (0.27°C [0.49°F]), but both devices were affected by environmental changes (−20 to 0 minutes) (Figure 1) and provided consistently lower measurements than RCT. Therefore, we conclude that oral temperature is an invalid measurement for estimating RCT in exercising individuals.

Axillary Temperature

Although some authors¹¹ have observed a strong correlation between axillary temperature and RCT in resting individuals, correlation does not imply validity.³³ Indeed, we found moderate correlations between AXL_{IE} and AXL_E and RCT (*r* = 0.77 and 0.60, respectively), but the mean bias for both devices was greater than the cutoff (−1.25°C [−2.24°F] and −0.94°C [−1.70°F] for AXL_{IE} and AXL_E, respectively). Kistemaker et al³⁴ also observed

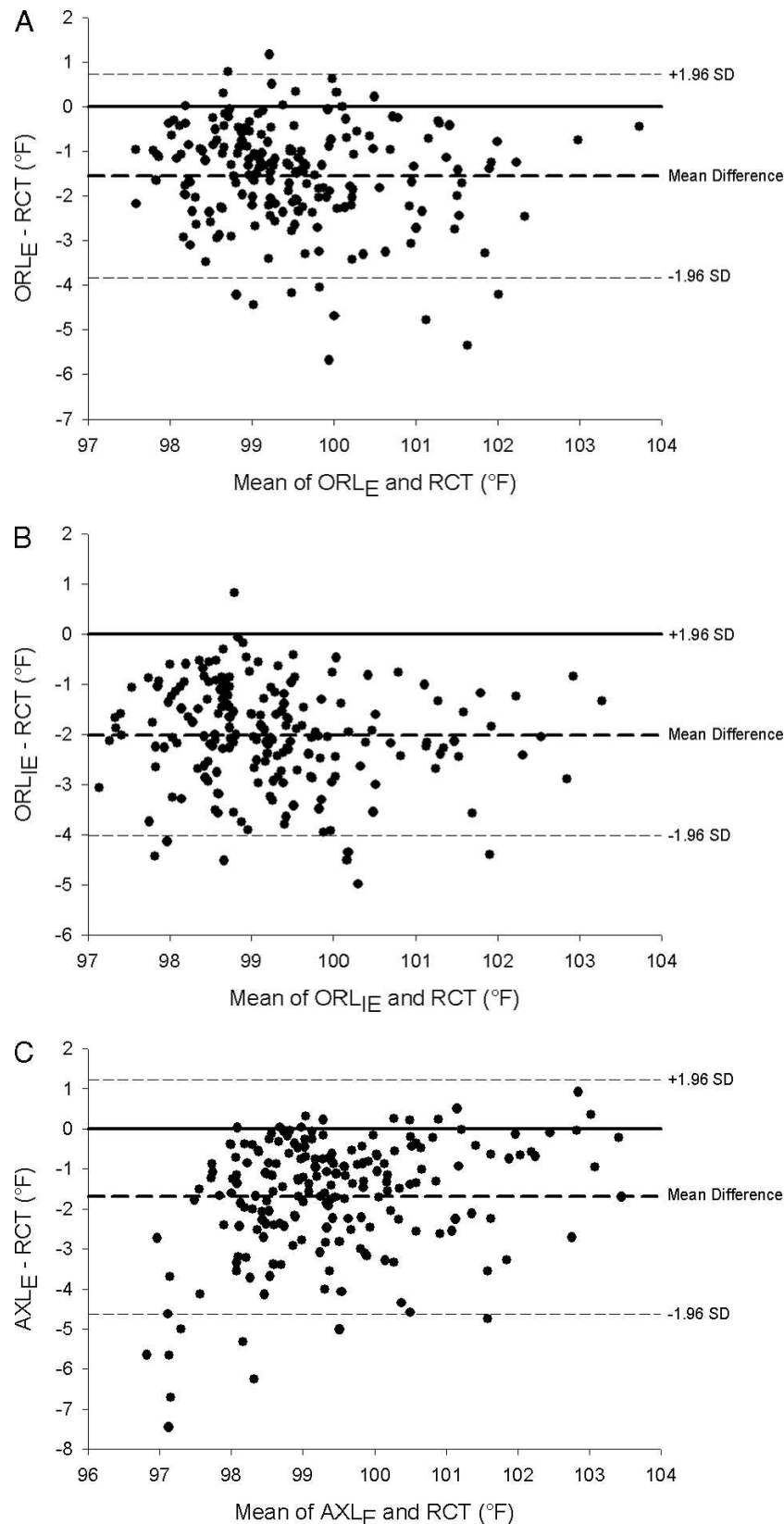


Figure 2. Bland-Altman plots indicating the mean bias (bold dashed line) and limits of agreement (dashed lines) for each temperature device compared with RCT. ORLE indicates oral temperature with expensive thermometer; ORLIE, oral temperature with inexpensive thermometer; AXLE, axillary temperature with expensive thermometer; AXLIE, axillary temperature with inexpensive thermometer; INT, intestinal temperature; AUR, aural temperature; TEM_{INST}, temporal temperature measured with the method described by the instructional manual; TEM_{MOD}, temporal temperature measured in a modified method; FST, forehead sticker temperature. (See text for further descriptions.)

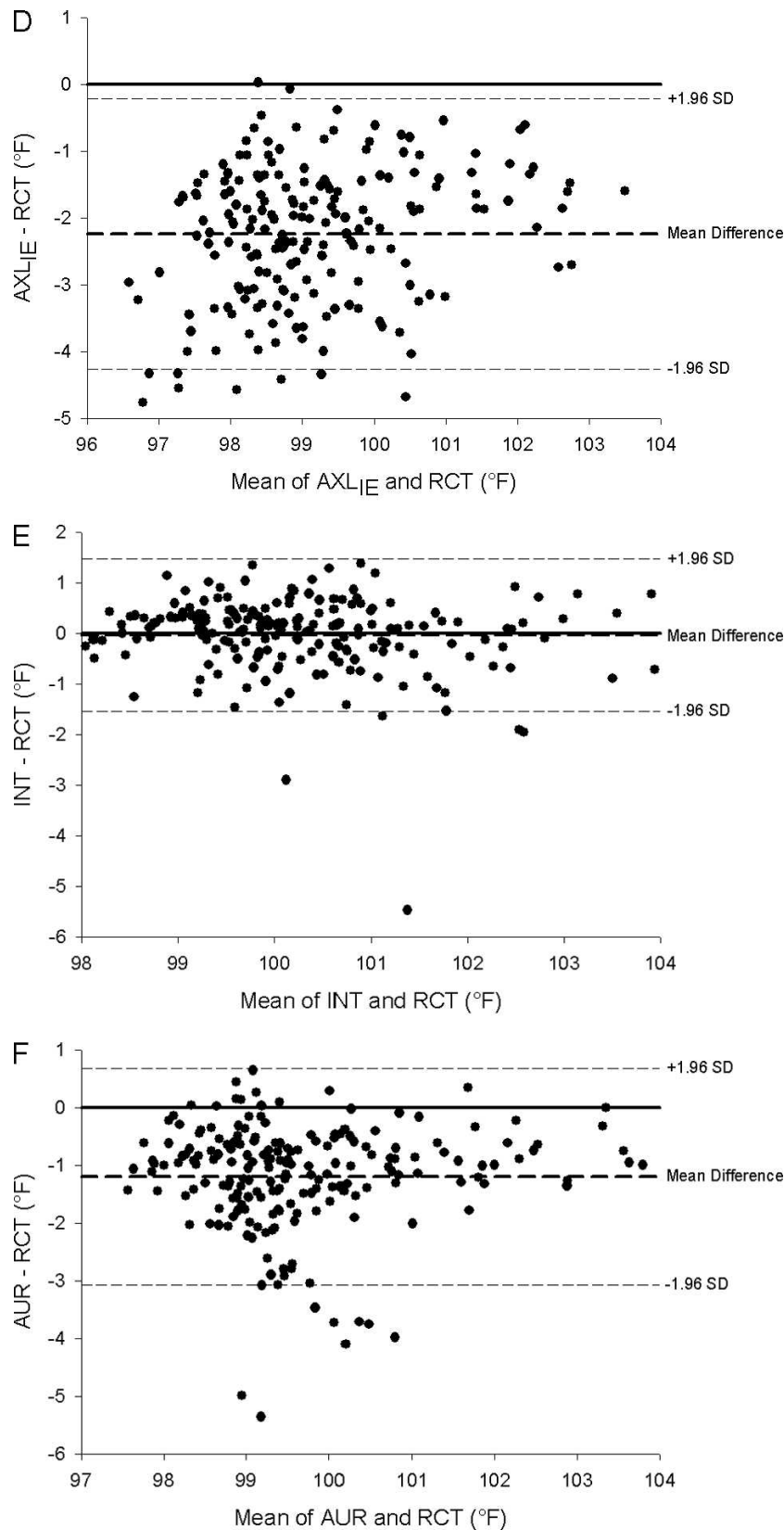


Figure 2. Continued.

that axillary temperature with indoor exercise failed to increase to the same degree as RCT.

We¹ have previously shown that axillary temperature during outdoor exercise is invalid. It appears that the mean bias of axillary versus RCT is greater outdoors (-2.07°C

$[-3.73^{\circ}\text{F}]$ and -2.58°C $[-4.65^{\circ}\text{F}]$ for AXLE and AXLE , respectively) than indoors in the present study (Table 1).¹ This finding may reflect the influence of uncontrollable environmental factors present outdoors (eg, wind, cloud cover). Axillary temperature measured by a standard

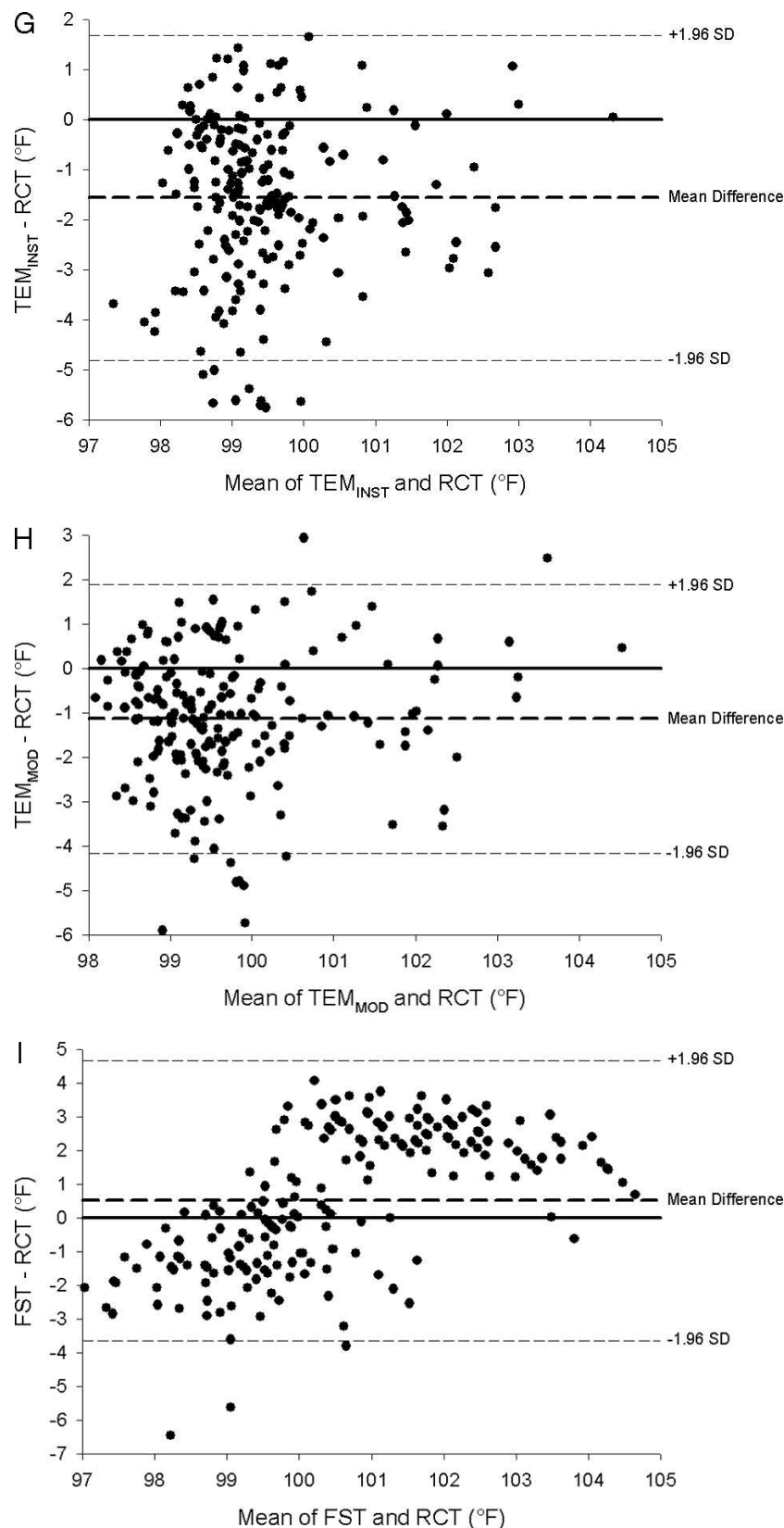


Figure 2. Continued.

temperature probe placed on the skin over the axillary artery often reflects skin temperature and not core temperature.⁸ Therefore, the higher temperature in this study (36.4°C [97.5°F]) versus outdoor exercise (29.7°C

[85.5°F])¹ may be a result of higher skin temperature and, thus, closer agreement between axillary temperature and RCT. This suggestion is supported by the increased mean difference between RCT and axillary temperature postex-

Table 2. Reliability of Temperature Devices That Were Taken Twice in Each Measurement Time Period

Temperature Measurement	Mean Difference	Pearson Correlation Coefficient, <i>r</i>	Limits of Agreement	Intraclass Correlation Coefficient	Standard Error of the Mean	Coefficient of Variation, (%) ^a
Intestinal temperature	0.01 ± 0.22°C 0.02 ± 0.40°F	0.94	±0.43°C ±0.78°F	0.97	±0.12°C ±0.21°F	1.80% (C) 1.22% (F)
Forehead temperature	0.03 ± 0.35°C 0.06 ± 0.64°F	0.97	±0.69°C ±1.25°F	0.98	±0.20°C ±0.36°F	3.69% (C) 2.52% (F)
Temporal measurement						
Per instruction manual	0.04 ± 0.33°C 0.10 ± 0.50°F	0.94	±0.65°C ±1.18°F	0.97	±0.14°C ±0.25°F	2.18% (C) 1.48% (F)
Modified method	0.09 ± 0.23°C 0.16 ± 0.42°F	0.96	±0.46°C ±0.82°F	0.98	±0.12°C ±0.21°F	2.21% (C) 1.50% (F)
Aural temperature	0.02 ± 0.19°C 0.04 ± 0.34°F	0.97	±0.37°C ±0.66°F	0.98	±0.11°C ±0.20°F	2.07% (C) 1.40% (F)

^a Value (C) = coefficient of variation when using °C; value (F) = coefficient of variation when using °F. See text for further descriptions.

ercise when participants were moved to a cooler environment (~23.3°C [74.0°F]; Figure 1).

Axillary temperature was lower than RCT at every time point during and after exercise in the heat. The mean biases between axillary (AXL_{IE} and AXL_E) temperature and RCT were greater than for the other devices tested. Although axillary temperature measured with a more expensive device (AXL_E) improved the agreement with RCT, both axillary devices (AXL_{IE} and AXL_E) were considered invalid. Therefore, we recommend that axillary temperature not be used to estimate internal body temperature in exercising individuals.

Intestinal Temperature

In numerous studies,^{1,22,28,35} INT, measured via a telemetric pill, has been shown to be a valid estimate of core body temperature during exercise. We found that INT versus other measures had the smallest bias, limits of agreement, standard error of the mean, and coefficient of variation; it also had the highest correlation (*r*) and intraclass correlation coefficient (Table 1). The mean bias between RCT and INT (−0.02°C [−0.03°F]) was lower than in other studies of indoor exercise (−0.15°C to 0.26°C).^{22,28} Differences may be due to different environmental temperatures, exercise protocols, and degrees of hyperthermia.

Telemetric pills have been used in outdoor settings to estimate core body temperature in professional athletes during practice³⁶ and competition.³⁷ According to the results of an outdoor validity study conducted by our laboratory,¹ INT was a valid estimate of internal body temperature. Mean bias was higher outdoors (−0.19°C [−0.34°F])¹ than indoors (−0.02°C [−0.03°F]), but these levels of bias are similar to those found in other studies^{22,28} and are still considered valid. The small bias we found may be due to the steady-state walking exercise.

The interday reliability of INT is very high (mean difference = 0.01°C).²⁸ Similarly, when repeated measures are taken within a temperature assessment time period, the reliability of INT is very high (mean difference = 0.02°C).³⁸ Our finding of a low mean difference (0.01°C [0.02°F]) between readings supports the reliability of telemetric pills.

Byrne and Lim,³⁹ after systematically reviewing studies comparing RCT, esophageal temperature, and INT, concluded that INT is a valid and reliable estimate of core

body temperature in a variety of settings, including outdoor exercise. Our data similarly show that INT was reliable and valid when assessing hyperthermia in individuals exercising indoors. Therefore, if circumstances allow for the ingestion of a telemetric pill at least 2 to 4 hours before exercise and retention of that pill during exercise, INT will be a valid alternative for measuring core body temperature.

Aural (Tympanic) Temperature

Although a direct measure of tympanic temperature (temperature probe touching the tympanic membrane) may be a valid estimation of core body temperature,⁴⁰ commercial devices attempt to estimate tympanic temperature by measuring infrared radiation from the tympanic membrane and then calculating temperature with a derived algorithm. The infrared radiation arises from the ear canal,⁴⁰ so these devices are measuring AUR, but this method often results in an invalid estimate of core body temperature.^{8,19,41,42}

Observing athletes exercising outdoors in the heat, we¹ previously found that AUR had a strong correlation with RCT (*r* = 0.70) but a large mean measurement difference (mean bias = −1.00°C [−1.80°F]). Similarly, with indoor exercise, the strong correlation (*r* = 0.77) does not represent the validity of AUR (mean bias = −0.67°C [−1.20°F]). Correlation coefficients do not determine validity.³³

In this study, AUR consistently underestimated RCT at every time point (mean bias = −0.67°C [−1.20°F]; Figure 1). This level of underestimation is consistent with that in other published studies (mean underestimation = 0.16°C to 1.07°C).^{1,8,42,43} Variations in AUR agreement with RCT can be attributed to differences in airflow,⁴² device used,⁸ and degree of hyperthermia⁴¹ and changes in skin temperature.¹⁹ Because all these factors may influence the ability of AUR to accurately estimate core body temperature, this method is not often recommended as a valid measure.^{1,19}

Temporal Temperature

Previous authors examining temporal TEM primarily studied infants at rest.^{5,6} Temporal temperature in this setting may⁵ or may not⁶ be valid. We found that at rest in mild air temperatures (23.3°C [74.0°F]), TEM was not different from RCT (*P* = .148 [TEM_{INST}] and .181

[TEM_{MOD}]). However, at rest and during and after exercise in a hot environment, TEM was different from RCT at each time point (Figure 1). Although the mean bias of TEM_{INST} (-0.87°C [-1.56°F]) improved slightly when using a modified technique (TEM_{MOD} mean bias = -0.63°C [-1.13°F]), neither method is a valid estimate of RCT.

Using the same device we used in the present study, Low et al³⁸ recently reported that temporal measurements failed to detect a 0.7°C increase in INT. They found that temporal temperature and INT were inversely related (slope = $-0.34 \pm 0.14^{\circ}\text{C}$) and poorly correlated ($R^2 = 0.29$). Other authors¹ measuring temporal temperature during outdoor exercise in the heat have observed an inverse relationship between RCT and temporal temperature. We also noted weak correlations between RCT and temporal temperature ($r = .25$ [TEM_{INST}] and $.38$ [TEM_{MOD}]); Table 1).

When compared with outdoor observations of TEM,¹ indoor exercise in a controlled environment lowered the degree of discrepancy between RCT and TEM. Mean bias outdoors was -1.46°C (-2.64°F) and -1.36°C (-2.44°F) for TEM_{INST} and TEM_{MOD}, respectively¹; in the present study, mean bias was -0.87°C (-1.56°F) and -0.63°C (-1.13°F) for TEM_{INST} and TEM_{MOD}, respectively. Kistemaker et al³⁴ examined TEM when participants cycled indoors for 30 minutes at 30°C . Temporal temperature underestimated internal body temperature at rest but overestimated it during exercise. On average, TEM overestimated core temperature by $0.50 \pm 0.50^{\circ}\text{C}$.³⁴ In the present study, RCT was underestimated during exercise (mean bias = -0.87 [TEM_{INST}] and -0.63°C [TEM_{MOD}]). The inability of TEM to accurately estimate RCT may be due to inconsistent blood flow in the superficial temporal artery,³⁴ different exercise durations, or different amounts of sweat present. Regardless, our data indicate that the measurement of TEM is an invalid estimate of RCT in exercising athletes.

Device reliability to measure TEM was assessed by repeating each measure during each temperature measuring period. The mean differences among subsequent readings were $0.04 \pm 0.33^{\circ}\text{C}$ ($0.10 \pm 0.50^{\circ}\text{F}$) and $0.09 \pm 0.23^{\circ}\text{C}$ ($0.16 \pm 0.42^{\circ}\text{F}$) for TEM_{INST} and TEM_{MOD}, respectively; (Table 2). Others³⁸ have measured greater variability among subsequent readings ($0.15^{\circ}\text{C} \pm 0.01$). Therefore, it appears that the device used to measure TEM in this study was reliable but not valid. Because a valid measure of RCT is the main objective when assessing an exercising individual for exertional heat stroke, we do not recommend TEM as an accurate estimate of RCT.

Liquid Crystal Forehead Strip

In resting individuals, forehead temperature measured with a liquid crystal forehead strip underestimated^{4,7} or overestimated⁴⁴ internal body temperature. With individuals exercising outdoors, we¹ previously found that in the shade (ie, absence of solar radiation), FST underestimated core body temperature, but on the field (ie, with solar radiation), FST overestimated RCT. This led to the conclusion that FST is influenced by environmental conditions and, thus, not recommended as a valid estimate of RCT.

With environmental conditions controlled indoors, we found that FST overestimated RCT during rest and exercise in the heat (mean difference = $1.28 \pm 0.48^{\circ}\text{C}$ [$2.30 \pm 0.86^{\circ}\text{F}$]). However, when before- and after-exercise measurements were taken in cooler ambient temperatures, FST underestimated RCT (mean difference = $-0.68 \pm 0.71^{\circ}\text{C}$ [$-1.22 \pm 1.28^{\circ}\text{F}$]; Figure 1). Because skin temperature often correlates with ambient conditions and because FST is measured in various ambient temperatures, FST is likely more influenced by changes in skin temperature than RCT.

When subsequent visual readings are acquired by the same person, FST shows strong reliability ($r = 0.97$) with small differences between readings (mean difference = $0.03 \pm 0.35^{\circ}\text{C}$ [$0.06 \pm 0.64^{\circ}\text{F}$]; Table 2). We do not know if the reliability is high between temperature assessors (interrater reliability).

Despite the influence of ambient temperature, mean bias of FST was 0.29°C (0.59°F) (Table 1). Although this bias was just beyond the designated cutoff for being considered a valid estimate of RCT, it should be noted that the bias depended on ambient temperature. Thus, FST is an invalid device to estimate internal body temperature in exercising individuals.

Thermal Sensation

In order to evaluate participants' thermal sensations, we used a visual scale adapted from Toner et al.²⁶ Ratings of thermal sensation weakly correlated with RCT ($r = 0.42$, $P < .001$). In the cooler environment before and after exercise, thermal sensation was 3.5 ± 0.8 , but it was not correlated with RCT ($r = -0.07$, $P = .497$). In the heat before and during exercise, thermal sensation was 5.8 ± 1.0 and was correlated with RCT ($r = 0.59$, $P < .001$). The differences in thermal sensation and RCT correlations are most likely due to rapid changes in skin temperature.²⁶

We¹ previously showed that thermal sensation during outdoor exercise was moderately correlated with RCT ($r = 0.72$). Therefore, it is evident that the strength of correlation between thermal sensation and RCT depends on the environment. Although thermal sensation may exhibit the same (increasing) trend as RCT, it is not a valid tool for estimating internal body temperature and should be avoided as a sole diagnostic tool for the evaluation of exertional heat stroke.

CONCLUSIONS

The purpose of our study was to evaluate the validity and reliability of commonly used devices to estimate core body temperature in hyperthermic individuals exercising indoors. Indoor exercise was performed in controlled laboratory conditions (ie, an environmental chamber). Future investigators should examine the validity and reliability of these temperature devices in other indoor settings and exercise modes.

Rectal temperature was used as the criterion standard because of its previously established validity,⁸⁻¹² ease of use, and practicability for the athletic trainer.² We conclude that measuring oral, forehead, aural, temporal, and axillary temperatures using the tested devices provided invalid measurements for the estimation of RCT. Although some devices may possess high reliability, the devices used

to measure these sites all had an average mean difference greater than the allowed cutoff for validity ($\pm 0.27^{\circ}\text{C}$ [$\pm 0.50^{\circ}\text{F}$]). The INT measured via a telemetric pill was the only measurement considered valid (mean bias = -0.02°C [-0.03°F]). Therefore, we recommend that the internal body temperature of exercising individuals be assessed with RCT. A previously ingested and retained intestinal telemetric pill is an acceptable alternative, but RCT must always remain a viable option in the event that measurement with telemetric pills is not possible (ie, malfunction, passed, not ingested, not enough time to pass through stomach).

Previous work in our laboratory¹ and our current findings reinforce the concept that temperature devices commonly used by medical professionals provide invalid estimates of core body temperature in indoor and outdoor settings. At this time, INT and RCT assessment are the only valid and reliable measurements for practical means of estimating core body temperature.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the researchers and technicians who made this study such a success: Chris Casa, Tutita Casa, Mike D'Alfonso, Christy Eason, Brian Gallagher, Ashleigh Gauvain, Neal Glaviano, Rob Huggins, Camille James, Nick Kalra, Jennifer Klau, Elaine Lee, Stephanie Mazerolle, Melissa Roti, Ian Scruggs, Barry Spiering, Kristin Stroly, Jakob Vingren, Greig Watson, Linda Yamamoto, and Brad Yeargin.

REFERENCES

1. Casa DJ, Becker SM, Ganio MS, et al. Validity of devices that assess body temperature during outdoor exercise in the heat. *J Athl Train*. 2007;42(3):333–342.
2. Binkley HM, Beckett J, Casa DJ, Kleiner DM, Plummer PE. National Athletic Trainers' Association position statement: Exertional heat illnesses. *J Athl Train*. 2002;37(3):329–343.
3. Dombek PM, Casa D, Yeargin SW, et al. Athletic trainers' knowledge and behavior regarding the prevention, recognition and treatment of exertional heat stroke at the high school level [abstract]. *J Athl Train*. 2006;41(2 suppl):S-47.
4. Kongpanichkul A, Bunjongpak S. A comparative study on accuracy of liquid crystal forehead, digital electronic axillary, infrared tympanic with glass-mercury rectal thermometer in infants and young children. *J Med Assoc Thai*. 2000;83(9):1068–1076.
5. Greenes DS, Fleisher GR. When body temperature changes, does rectal temperature lag? *J Pediatr*. 2004;144(6):824–826.
6. Greenes DS, Fleisher GR. Accuracy of a noninvasive temporal artery thermometer for use in infants. *Arch Pediatr Adolesc Med*. 2001;155(3):376–381.
7. Allen GC, Horrow JC, Rosenberg H. Does forehead liquid crystal temperature accurately reflect "core" temperature? *Can J Anaesth*. 1990;37(6):659–662.
8. Jensen BN, Jensen FS, Madsen SN, Lossel K. Accuracy of digital tympanic, oral, axillary, and rectal thermometers compared with standard rectal mercury thermometers. *Eur J Surg*. 2000;166(11):848–851.
9. Lee SM, Williams WJ, Fortney Schneider SM. Core temperature measurement during supine exercise: esophageal, rectal, and intestinal temperatures. *Aviat Space Environ Med*. 2000;71(9):939–945.
10. Lefrant JY, Muller L, de La Coussaye JE, et al. Temperature measurement in intensive care patients: Comparison of urinary bladder, oesophageal, rectal, axillary, and inguinal methods versus pulmonary artery core method. *Intensive Care Med*. 2003;29(3):414–418.
11. Chaturvedi D, Vilhekar KY, Chaturvedi P, Bharambe MS. Comparison of axillary temperature with rectal or oral temperature and determination of optimum placement time in children. *Indian Pediatr*. 2004;41(6):600–603.
12. Casa DJ, Armstrong LE, Ganio MS, Yeargin SW. Exertional heat stroke in competitive athletes. *Curr Sports Med Rep*. 2005;4(6):309–317.
13. Roberts WO. Assessing core temperature in collapsed athletes: what's the best method? *Physician Sportsmed*. 1994;22(8):49–55.
14. Moran DS, Mendall L. Core temperature measurement: methods and current insights. *Sports Med*. 2002;32(14):879–885.
15. Casa DJ, Roberts WO. Considerations for the medical staff: Preventing, identifying, and treating exertional heat illnesses. In: Armstrong LE, ed. *Exertional Heat Illnesses*. Champaign, IL: Human Kinetics; 2003:169–196.
16. Brown GA, Williams GM. The effect of head cooling on deep body temperature and thermal comfort in man. *Aviat Space Environ Med*. 1982;53(6):583–586.
17. Livingstone SD, Grayson J, Frim J, Allen CL, Limmer RE. Effect of cold exposure on various sites of core temperature measurements. *J Appl Physiol*. 1983;54(4):1025–1031.
18. Zehner WJ, Terndrup TE. The impact of moderate ambient temperature variance on the relationship between oral, rectal, and tympanic membrane temperatures. *Clin Pediatr (Phila)*. 1991;30(4 suppl):61–72.
19. Deschamps A, Levy RD, Cosio MG, Marliss EB, Magder S. Tympanic temperature should not be used to assess exercise induced hyperthermia. *Clin J Sport Med*. 1992;2(1):27–32.
20. Cabanac M, Caputa M. Natural selective cooling of the human brain: evidence of its occurrence and magnitude. *J Physiol*. 1979;286(1):255–264.
21. Shiraki K, Sagawa S, Tajima F, Yokota A, Hashimoto M, Brengelmann GL. Independence of brain and tympanic temperatures in an anesthetized human. *J Appl Physiol*. 1988;65(1):482–486.
22. Kolka MA, Quigley MD, Blanchard LA, Toyota DA, Stephenson LA. Validation of a temperature telemetry system during moderate and strenuous exercise. *J Therm Biol*. 1993;18(4):203–210.
23. Jackson AS, Pollock ML. Practical assessment of body composition. *Physician Sportsmed*. 1985;13(5):76–90.
24. Pollock ML, Schmidt DH, Jackson AS. Measurement of cardiorespiratory fitness and body composition in the clinical setting. *Compr Ther*. 1980;6(9):12–27.
25. Heyward VH, Stolarczyk LM. *Applied Body Composition Assessment*. Champaign, IL: Human Kinetics; 1996:12.
26. Toner MM, Drolet LL, Pandolf KB. Perceptual and physiological responses during exercise in cool and cold water. *Percept Mot Skills*. 1986;62(1):211–220.
27. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26(4):217–238.
28. 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.
29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
30. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations, part I: correlation within subjects. *BMJ*. 1995;310(6977):446.
31. Mairiaux P, Sagot JC, Candau V. Oral temperature as an index of core temperature during heat transients. *Eur J Appl Physiol Occup Physiol*. 1983;50(3):331–341.
32. Edwards RJ, Belyavin AJ, Harrison MH. Core temperature measurement in man. *Aviat Space Environ Med*. 1978;49(11):1289–1294.
33. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135–160.
34. Kistemaker JA, Den Hartog EA, Daanen HA. Reliability of an infrared forehead skin thermometer for core temperature measurements. *J Med Eng Technol*. 2006;30(4):252–261.
35. Kolka MA, Levine L, Stephenson LA. Use of an ingestible telemetry system to measure core temperature under chemical protective clothing. *J Therm Biol*. 1997;22(4–5):343–349.
36. Godek SF, Bartolozzi AR, Burkholder R, Sugarman E, Dorshimer G. Core temperature and percentage of dehydration in professional football linemen and backs during preseason practices. *J Athl Train*. 2006;41(1):8–17.

37. Edwards AM, Clark NA. Thermoregulatory observations in soccer match play: professional and recreational level applications using an intestinal pill system to measure core temperature. *Br J Sports Med.* 2006;40(2):133–138.
38. Low DA, Vu A, Brown M, et al. Temporal thermometry fails to track body core temperature during heat stress. *Med Sci Sports Exerc.* 2007;39(7):1029–1035.
39. 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.
40. Sato KT, Kane NL, Soos G, Gisolfi CV, Kondo N, Sato K. Reexamination of tympanic membrane temperature as a core temperature. *J Appl Physiol.* 1996;80(4):1233–1239.
41. Armstrong LE, Maresh CM, Crago AE, Adams R, Roberts WO. Interpretation of aural temperatures during exercise, hyperthermia, and cooling therapy. *Med Exerc Nutr Health.* 1994;3(1):9–16.
42. Hansen RD, Daley WH, Leelarthapin B. The effect of facial airflow on the estimation of exercise core temperature by infrared tympanic thermometry. *Aust J Sci Med Sport.* 1993;25(1):26–31.
43. Roth RN, Verdile VP, Grollman LJ, Stone DA. Agreement between rectal and tympanic membrane temperatures in marathon runners. *Ann Emerg Med.* 1996;28(4):414–417.
44. Patel N, Smith CE, Pinchak AC, Hagen JF. Comparison of esophageal, tympanic, and forehead skin temperatures in adult patients. *J Clin Anesth.* 1996;8(6):462–468.

Matthew S. Ganio, MS, contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Christopher M. Brown, MA, ATC, contributed to conception and design, acquisition and analysis and interpretation of the data, and critical revision and final approval of the article. Douglas J. Casa, PhD, ATC, FNATA, FACSM, contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Shannon M. Becker, MA, ATC, contributed to conception and design, acquisition and analysis and interpretation of the data, and critical revision and final approval of the article. Susan W. Yeargin, PhD, ATC; Brendon P. McDermott, MS, ATC; and Lindsay M. Boots, BS, ATC, contributed to conception and design, acquisition of the data, and critical revision and final approval of the article. Paul W. Boyd, BS, ATC, contributed to acquisition of the data and critical revision and final approval of the article. Lawrence E. Armstrong, PhD, FACSM, and Carl M. Maresh, PhD, FACSM, contributed to conception and design, acquisition of the data, and critical revision and final approval of the article.

Address correspondence to Matthew S. Ganio, MS, University of Connecticut, 2095 Hillside Road, U-1110, Storrs, CT 06269. Address e-mail to matthew.ganio@uconn.edu.