BRIEF REPORT

Homemade Thermometry Instruments in the Field

Mathieu Pasquier, MD; Valentin Rousson, Pr; Grégoire Zen Ruffinen, MD; Olivier Hugli, MD, MPH

From the Emergency Service, University Hospital Center, Lausanne, Switzerland (Drs Pasquier and Hugli); Air-Glacières, Maison François-Xavier Bagnoud du Sauvetage, GRIMM, Sion, Switzerland (Drs Pasquier and Zen Ruffinen); and the Institute for Social and Preventive Medicine, Statistical Unit, University Hospital Center, Lausanne, Switzerland (Dr Rousson).

Objective.—Esophageal temperature is the gold standard for in-the-field temperature monitoring in hypothermic victims with cardiac arrest. For practical reasons, some mountain rescue teams use homemade esophageal thermometers to measure esophageal temperature; these consist of nonmedical inside/outside temperature monitoring instruments that have been modified to allow for esophageal insertion. We planned a study to determine the accuracy of such thermometers.

Methods.—Two of the same model of digital cabled indoor/outdoor thermometer were modified and tested in comparison with a reference thermometer. The thermometers were tested in a water bath at different temperatures between 10°C and 35.2°C. Three hundred measurements were taken with each thermometer.

Results.—Our experimental study showed that both homemade thermometers provided a good correlation and a clinically acceptable agreement in comparison with the reference thermometer. Measurements were within 0.5°C in comparison with the reference thermometer 97.5% of the time.

Conclusions.—The homemade thermometers performed well in vitro, in comparison with a reference thermometer. However, because these devices in their original form are not designed for clinical use, their use should be restricted to situations when the use of a conventional esophageal thermometer is impossible.

Key words: hypothermia, temperature, avalanche, mountain medicine, thermometer

Introduction

Esophageal temperature is the gold standard for in-the-field temperature monitoring in hypothermic victims with cardiac arrest.^{1,2} In the specific case of avalanche victims, esophageal temperature provides key data for deciding whether or not to continue or to withhold resuscitative efforts in asystolic patients; a cutoff of 32°C is currently used.^{2,3} Several rescue teams in North America and Europe use homemade esophageal thermometers in the field. These consist of nonmedical inside/outside temperature monitoring instruments that have been modified to allow for esophageal insertion. The main advantages of these devices, apart from their low cost, are that they are compact and light and, therefore, ideally adapted for on-foot rescue missions as well as high altitude or

wilderness operations. The accuracy of these devices has not, however, been evaluated to date. We sought to experimentally evaluate the reliability and accuracy of such thermometers in comparison with the temperature measurements of a standard reference thermometer.

Methods

Two of the same model of digital cabled indoor/outdoor thermometer (TFA 30.2018-02, TFA-Dostmann GmbH & CO KG, Wertheim-Reicholzheim, Germany) were tested (named TFA1 and TFA2). The probes were modified through reinforcement with a metallic wire that was then covered with heat-shrink tubing as shown in Figure 1. The reference temperature was determined using the G100 thermometer (Geotechnical Instruments Ltd, Warwickshire, UK). Immediately before the study, calibration of this reference thermometer was verified by the manufacturer. The accuracy of the reference thermometer is $\pm 0.2^\circ\text{C}$ from 32°C to 44°C, and $\pm 0.5^\circ\text{C}$ over the rest of the range of 0°C to +50.0°C.

The authors have no conflicts of interest to declare, and there are no sources of support or funding to declare.

Corresponding author: Mathieu Pasquier, MD, Emergency Service, University Hospital Center Lausanne, 1011 Lausanne, Switzerland (e-mail: mathieu.pasquier@chuv.ch).

Figure 1. The TFA 30.2018-02 thermometer costs about US \$17. It measures $39 \times 52 \times 15$ mm, weighs 50 g, and functions with a 1.5-V button cell battery. The modified probe is shown on the right.

The thermometers were tested simultaneously in the same water bath at temperatures between 10°C and 35.2°C . Each measurement period consisted of simultaneous recordings of the 3 thermometers, whose probes were inserted to the same depth in the water. At least 1 minute was allowed after the probe was inserted for thermal equilibration. Hot or cold water was then mixed in the water bath to obtain different water temperatures between 10°C and 35.2°C , and subsequent measurements were made until a total of 300 measurements were recorded for each of the 3 probes. The temperature increments were about 0.2°C to 0.4°C , depending on the volume of hot or temperate water that was added. All measurements were carried out by one experienced researcher who was trained to use the devices. Another researcher recorded and transcribed the temperature measurements in the database in blinded fashion.

STATISTICAL METHODS

Agreement between the gold standard (the reference thermometer) and the measurements made with either TFA1 or TFA2 was assessed using both a concordance correlation,⁴ which is equivalent to an intraclass correlation coefficient, and limits of agreement.⁵ Contrary to the classical Pearson correlation, concordance/intraclass correlation penalizes the systematic error between 2 measurements, and provides information on the percentage of variance in the measurements that is not attributable to measurement errors. On the other hand, limits of agreement calculated according to Bland and Altman⁵ provide information on the size of the discrepancies

between 2 measurements, allowing one to judge whether they are clinically acceptable or not. Clinically acceptable limits of agreement were a priori defined to be $\pm 0.5^{\circ}\text{C}$.^{6,7} We considered the experimental thermometers to be equivalent if the estimated 95% limits of agreement with the reference thermometer reading were within 0.5°C , that is, if 95% of measurements were within 0.5°C of the reference reading.⁶ However, one of the assumptions when using such limits of agreements is that the mean difference between 2 measurements is constant throughout the range of measurements,⁸ which might not be the case in our study. We, therefore, added a regression model with the gold standard as the response variable and either TFA1 or TFA2 as the predictor. This allowed us to calculate 95% prediction intervals for the gold standard depending on the various temperatures measured by TFA1 or TFA2. As mentioned by Bland and Altman,⁸ “this gives us something akin to the limits of agreement,” while allowing us to describe a nonconstant mean difference. The statistical analysis and the graphics were completed using the free statistical software R (version 2.5.1; Available at <http://www.r-project.org/>).

Results

Three hundred measurements were taken with each thermometer. Temperatures ranged from 10°C to 35.2°C for the reference thermometer.

As shown in the top panels of Figure 2, the 3 measurements (gold standard, TFA1, TFA2) were highly correlated as the concordance correlation between the gold standard and either TFA1 or TFA2 was greater than 0.99. From a clinical perspective, it is remarkable to note that 95.5% of the measurements made with either TFA1 or TFA2 were found to be within 0.5°C of the gold standard. Fifteen measurements made with TFA2 differed by 0.6°C with the gold standard, and all at low reference temperatures of between 11.2°C and 13.3°C .

On average, measurements made with either TFA1 or TFA2 were found to be slightly below the gold standard. The average of the differences between TFA1 and the gold standard was -0.127°C with a standard deviation of 0.162°C . The average of the differences between TFA2 and the gold standard was -0.259°C with a standard deviation of 0.167°C . Thus, approximately 95% of the differences between TFA1 and the gold standard were between $-0.127^{\circ}\text{C} \pm 1.96 \times 0.162^{\circ}\text{C}$, ie, between -0.44°C and $+0.19^{\circ}\text{C}$, while approximately 95% of the differences between TFA2 and the gold standard were between $-0.259^{\circ}\text{C} \pm 1.96 \times 0.167^{\circ}\text{C}$, ie, between -0.59°C and $+0.07^{\circ}\text{C}$. This corresponds to the limits of agreement calculated according to Bland and Altman,⁵ which are shown in the bottom panels of Figure 2.

Figure 2. Top panels show scatterplots of measurements made with TFA1 (top left panel) and TFA2 (top right panel) versus the gold standard. A dotted identity line is plotted as a reference line. Bottom panels show the differences of measurements (bottom left panel: TFA1 – gold standard; bottom right panel: TFA2 – gold standard) versus the average of measurements, together with the limits of agreement calculated according to Bland and Altman.⁵ A dotted horizontal line at zero is plotted as a reference line.

Note, however, that the clear positive relationships in these graphics suggest that the discrepancy between TFA1 or TFA2 and the gold standard was on average more negative at low temperatures than at high temperatures (the correlation between the average and the difference of the two measurements being 0.84 for TFA1 and 0.76 for TFA2, $P < .0001$). To describe this trend, we calculated 95% prediction intervals for what would be the gold standard given the various temperatures measured by TFA1 or TFA2 using a classic regression model. We found that the measurement of the gold standard would lie between 10.21°C and 10.56°C with a probability of 95% when TFA1 is equal to 10°C, and between 34.96°C and 35.31°C when it is equal to 35.2°C. Thus, prediction intervals for the gold standard were within the limits defined by $\text{TFA1} \pm 0.5^\circ\text{C}$ for high measurements of TFA1, whereas they were slightly outside these limits for low measurements of TFA1. In fact, prediction intervals for the gold standard were within the limits defined by $\text{TFA1} \pm 0.5^\circ\text{C}$ for all measurements of

TFA1 that were above 12.9°C. On the other hand, we found that the measurement of the gold standard would lie between 10.29°C and 10.71°C with a probability of 95% when the measurement of TFA2 is equal to 10°C, and between 35.07°C and 35.49°C when it is equal to 35.2°C. Prediction intervals for the gold standard were within the limits defined by $\text{TFA2} \pm 0.5^\circ\text{C}$ for all measurements of TFA2 above 22.4°C.

Discussion

Our experimental study showed that both homemade thermometers provided a good correlation and a clinically acceptable agreement in comparison with the reference thermometer. Measurements were within 0.5°C in comparison with the reference thermometer 97.5% of the time. The discrepancy between the 2 homemade thermometers was found to be more negative at very low temperatures. The differences were, as it was a priori defined, considered as clinically acceptable at tempera-

tures above 12.9°C for TFA1, and at temperature above 22.4°C for TFA2.⁶

A traditional statistical approach like correlation coefficients may provide an inadequate measure to validate a new measurement instrument.⁹ We therefore used the Bland and Altman method to determine the agreement between the different thermometers.⁵ This statistical method is generally accepted as the most appropriate to compare methods measuring the same clinical parameter.⁹ The Bland and Altman graph plots the difference in measurements between the 2 methods against the mean of those measurements. It can be interpreted as follows: if the values that fall within 2 standard deviations of the mean difference are not clinically significant (ie, if the limits of agreement are small), the 2 methods are considered to be in agreement and the new method can be judged acceptable.⁹ When the instrument bias, which is determined by the mean difference between the experimental and the reference thermometer, is close to zero, the 2 methods agree.⁵

Ultimately, any new medical device should undergo clinical evaluation in an *in vivo* setting. A major barrier for human testing of these homemade thermometers rests in the fact that it is difficult to develop homologated materials for human research and to test them for biocompatibility. On the other hand, *in vitro* testing offers obvious advantages. It allows a greater number of measures than *in vivo* testing, and is much less expensive. Water bath calibration has been chosen because it is considered the gold standard calibration procedure for temperature measuring instruments.¹⁰ It also enables testing the device in a wide range of temperatures, such as those encountered in the field, in contrary to clinical studies that are usually characterized by the absence of markedly hypothermic patients.⁷ This is of major importance as those types of homemade thermometers are mostly used in the context of avalanche hypothermia victims.

How, then, is it possible that physicians can use such low-tech methods of temperature monitoring in the 21st century? There are 2 main reasons that would bring some to find this method unacceptable. The first objection is that the materials used are not tested for biocompatibility or for clinical tolerance. The second is that these thermometers are not certified to provide reliable values with acceptable accuracy. Despite being aware of these facts, many mountain or remote area rescue units in North America and Europe nevertheless use such inexpensive modified indoor/outdoor thermometers during selected rescue missions. Prehospital rescue teams regularly use multiparameter monitors with esophageal probes for temperature measurement. In some circumstances, however, their use is extremely impractical. Examples include rescues on foot or highly technical rescue missions,

including medicine in remote environments. In those settings, it is often illusory to bring along conventional temperature monitoring devices because of weight and volume constraints as the medical materials have to be carried in addition to the personal and safety equipment. This is probably the main reason why this type of homemade thermometer has been developed by rescue teams as a “better than nothing” option. Finally, it must be noted that laboratory thermometers with a stiff metallic probe, such as those used for rectal temperature for on-site forensic purposes, are not suitable for emergency medical use, as perforation of the rectum is a common side effect.

The core temperature is of great concern in mountain medicine, as timely management decisions depend on it to guide the course of immediate treatment. In avalanche victims, a core (ie, esophageal) temperature of 32°C is used for triaging patients in asystolic cardiac arrest.^{2,3} Nearly one third of the reference temperatures we measured were around this 32°C key temperature (between 30° and 34°C). All measurements provided by the homemade thermometers in this temperature range were within $\pm 0.5^\circ\text{C}$ of the reference thermometer reading. The greatest difference between the homemade thermometers and the gold standard was 0.6°C, and this occurred at reference temperatures between 11.2°C and 13.3°C, ie, below the 13.7°C that is the lowest reversible core temperature reported to date in case of accidental hypothermia.

These satisfactory results should be interpreted in conjunction with the numerous limitations of these thermometers. If the problem of the biocompatibility can be reasonably set aside when talking about patients in cardiac arrest, some important concerns nevertheless persist. The main one is the influence of the environment on the accuracy of the measurement. If the thermometer casing is not at the normal room temperature, as it was in our study, it is known that electronic thermometers may be less accurate.¹¹ This limitation, which also applies to conventional medical thermometers, could be overcome in the field by using a heated case, or keeping the thermometer case within the rescuer's jacket, in close contact with his or her body. Another limitation is the absence of medical homologation of these thermometers, ie, the absence of accuracy certification and quality control. However, regardless of what might be done to improve this kind of technique, treatments and decisions as crucial as withholding or pursuing resuscitation efforts on a patient should under no circumstances be based on a single temperature measurement with one of these homemade thermometers. Future research from the industry would be welcome to develop a medical thermometer that would be validated at low ambient temper-

atures, and also be light and inexpensive. An interesting line of research would be to develop a compact thermometer that would be compatible with the current medical esophageal probes. However, because of the few potential buyers, interest from the industry for allocating resources to such development is probably limited.

The main limitation of our study is that we only evaluated two identical copies of one model of thermometer. We cannot extrapolate our results to other models or generalize them to other copies of this same model of thermometer. However, our study provides the methodology that can be used by any rescuer to test the accuracy of an individual thermometer (R software commands used for our validation are available from the corresponding author).

Conclusions

The homemade thermometers modified to allow for esophageal measurement to estimate core temperature performed well in vitro, in comparison with a reference thermometer. However, because of major concerns in terms of biocompatibility and quality control of these devices that are not intended for clinical use, their use should be restricted to situations when the use of a conventional esophageal thermometer is impossible. Critical clinical decisions should not be based purely on such homemade thermometry instrument readings: rather, the measurements should be considered in a global context along with other medical information. Because the accuracy of core temperature monitoring in the field fronts numerous practical limits, its use as a means of triaging asystolic avalanche victims remains debatable.

Acknowledgments

We thank Grégoire Carron for his help with the modification of the thermometers and Fabienne Pourchet for

her help with data collection. We also thank Danielle Wyss for proofreading and final translation.

References

1. Durrer B, Brugger H, Syme D; International Commission for Mountain Emergency Medicine. The medical on-site treatment of hypothermia: ICAR-MEDCOM recommendation. *High Alt Med Biol.* 2003;4:99–103.
2. Brugger H, Durrer B, Adler-Kastner L, Falk M, Tschirky F. Field management of avalanche victims. *Resuscitation.* 2001;51:7–15.
3. Boyd J, Brugger H, Shuster M. Prognostic factors in avalanche resuscitation: a systematic review. *Resuscitation.* 2010;81:645–652.
4. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989;45:255–268.
5. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–310.
6. Langham GE, Maheshwari A, Contrera K, You J, Mascha E, Sessler DI. Noninvasive temperature monitoring in post-anesthesia care units. *Anesthesiology.* 2009;111:90–96.
7. Giuliano KK, Scott SS, Elliot S, Giuliano AJ. Temperature measurement in critically ill orally intubated adults: a comparison of pulmonary artery core, tympanic, and oral methods. *Crit Care Med.* 1999;27:2188–2193.
8. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol.* 2003;22:85–93.
9. Lenhardt R. Monitoring and thermal management. *Best Pract Res Clin Anaesthesiol.* 2003;17:569–581.
10. Hooper VD, Andrews JO. Accuracy of noninvasive core temperature measurement in acutely ill adults: the state of the science. *Biol Res Nurs.* 2006;8:24–34.
11. Crawford DC, Hicks B, Thompson MJ. Which thermometer? Factors influencing best choice for intermittent clinical temperature assessment. *J Med Eng Technol.* 2006;30:199–211.