Identifying Critical Design Parameters for Improved Body Temperature Measurements: A Clinical Study Comparing Transient and Predicted Temperature Measurements

Oleg Vesnovsky[1,2]
FDA Center for Devices & Radiological Health,
Silver Spring, MD 20993
e-mail: oleg.vesnovsky@fda.hhs.gov

Liang Zhu[1]
Department of Mechanical Engineering,
University of Maryland,
Baltimore County,
Baltimore, MD 21250

Laurence W. Grossman
FDA Center for Devices & Radiological Health,
Silver Spring, MD 20993

Jon P. Casamento
FDA Center for Devices & Radiological Health,
Silver Spring, MD 20993

Alireza Chamani
Department of Mechanical Engineering,
University of Maryland,
Baltimore County,
Baltimore, MD 21250

Nadeesri Wijekoon
Department of Mathematics and Statistics,
University of Maryland,
Baltimore County,
Baltimore, MD 21250

L. D. Timmie Topoleski[2]
FDA Center for Devices & Radiological Health,
Silver Spring, MD 20993;
Department of Mechanical Engineering,
University of Maryland,
Baltimore County,
Baltimore, MD 21250
e-mail: topoleski@umbc.edu

Readily available store brand, or "home," thermometers are used countless times in the home and clinic as a first diagnostic measure of body temperature. Measurement inaccuracies may lead to unnecessary medical visits or medication (false positives), or, potentially worse, lack of intervention when a person is truly sick (false negatives). A critical first step in the design process is to determine the shortcomings of the existing designs. For this project, we evaluated the accuracy of three currently available store brand thermometers in a pediatric population. The accuracies of the thermometers were assessed by comparing their body temperature predictions to those measured by a specially designed and calibrated and fast-responding reference thermometer. The reference thermometer was placed at the measurement site simultaneously with the store brand thermometer and recorded the temperature at the measurement site continuously. More than 300 healthy or sick pediatric subjects were enrolled in this study. Temperatures were measured at both the oral and axillary (under the arm) sites. The store brand thermometer measurements characteristically deviated from the reference thermometer temperature after 120 s, and the deviations did not follow a consistent pattern. The Brand C thermometers had the greatest deviations of up to 3.7 °F (2.1 °C), while the Brand A thermometers had the lowest deviations; however, they still deviated by up to 1.9 °F (1.1 °C). The data showed that the tested store brand thermometers had lower accuracy than the ±0.2 °F (0.1 °C) indicated in their Instructions for Use. Our recorded reference (transient) data showed that there was a wide variation in the transient temperature profiles. The store brand thermometers tested stated in their documentation that they are able to predict a body temperature based on transient temperature values over the first 5–10 s of measurements, implying that they use an embedded algorithm to extrapolate to the steady-state temperature. Significant deviations from the maximum temperature after time $t = 4.6t_{0.63}$ illustrated that the transient temperature profiles may not be represented by an exponential function with a single time constant, $t_{0.63}$. The accuracy of those embedded algorithms was not confirmed by our study, since the predicted body temperatures do not capture the large variations observed over the initial 10 s of the measurements. A thermometer with an error of several degrees Fahrenheit may result in a false positive or negative diagnosis of fever in children. The transient temperature measurements from our clinical study represent unique and critical data for helping to design the next generation of readily available, highly accurate, home thermometers.
[DOI: 10.1115/1.4041589]

Keywords: bioheat transfer, body temperature, store brand thermometer, accuracy, clinical study

1 Introduction

Since the invention of the clinical mercury thermometer by Sir Thomas Clifford Allbutt in 1866, mercury-in-glass thermometers have been widely used in clinical settings to measure body temperatures [1]. Body temperature monitoring is an important tool for diagnosing infections, detecting fever, monitoring thermoregulation during surgery, and assessing postsurgical recovery [2–4]. Readily available store brand, or home, thermometers are used countless times in the home and clinic as a first diagnostic measure of body temperature. Measurement inaccuracies may lead to unnecessary medical visits or medication (false positives), or, potentially worse, lack of intervention when a person is truly sick (false negatives).

The central control for thermoregulation is located in the hypothalamus. Since the hypothalamus is not easily accessible by thermometers, other body locations have been identified as alternative measuring sites. Different sites used to measure body temperature include the pulmonary artery, rectum, bladder, distal esophagus and nasopharynx, under the tongue, under the armpit, tympanic

[1]O. Vesnovsky and L. Zhu both contributed equally.
[2]Corresponding authors.

Manuscript received May 16, 2017; final manuscript received September 24, 2018; published online December 4, 2018. Assoc. Editor: Xiaoming He.

membrane, and forehead [5–8]. Common sites for routine temperature measurements at home or in a clinical setting are the underside of the tongue, the forehead, the ear, and the rectum.

Thermometer design depends on the measurement site for that thermometer; for example, thermometers used orally have a different mechanism for measuring temperature than thermometers used in the ear or on the forehead. Another design consideration has been in the choice of materials used for thermometers; for example, because of mercury's known toxicity, other liquids, such as alcohol and Galinstan®, have replaced mercury [9].

Further, as body temperature measurement technology has advanced, different analog and digital thermometers have been developed to provide noninvasive, or minimally invasive, temperature measurements, to shorten the measurement time, and to provide high measurement precision. Unlike traditional mercury thermometers that require at least several minutes to establish thermal equilibrium, the new technologies incorporate thermistor beads or thermocouples to convert the measured resistance or voltage into temperature readings (referred to as electronic thermometers), or infra-red sensors. The small sizes of the thermocouples, thermistor beads, or the infra-red sensors result in a much shorter time to establish a temperature measurement. They also facilitate digital display of the temperature readings.

The temperatures measured under the armpit, the tympanic membrane, and the forehead may not represent the true body core temperature, even after a steady-state measurement is established. Thermometers used in those three sites often require rigorous calibrations with other sites to determine their relationships with the true body core temperature. Some commercially available thermometers have been developed to predict true body core temperatures by extrapolating from values taken during the first few seconds of measurement.

Thus, there are two major sources of potential inaccuracies in electronic thermometers. One source is the deviation from the true body core temperature at the measurement site chosen for convenience, even after a steady-state temperature field is established. The second source arises from reducing the time required to produce a temperature measurement, also motivated by convenience.

Axillary temperatures are less accurate because their measurement locations are away from the body core. Lawson et al., in an experiment where temperatures were measured in the pulmonary artery or the rectum, showed that the axillary temperatures were lower than the core by 0.2 °C for normal subjects and 0.6 °C for hyperthermic subjects [8]. A study by Erickson and Meyer also showed that the axillary temperature readings were lower than those made by a pulmonary thermometer by approximately 0.6 °C [5]. In two other studies, temperatures measured by digital axillary thermometers were less than core body temperatures measured by rectal thermometers by approximately 0.28 °C [10], and 0.64 °C [11]. Kara et al. also used axillary sites to compare a digital thermometer to a mercury-in-glass thermometer and found discrepancies in temperature measurements from +0.37 °C to −1.24 °C. In addition, the digital thermometer misdiagnosed 20% of febrile children as afebrile [12]. The reported research suggests that axillary temperature measurements must include some correction factors, like a positive offset, to provide an accurate body temperature reading.

Many digital thermometers use an undisclosed algorithm to predict the equilibrium temperature based on the readings made during only the initial seconds of measurement (the "predictive mode") [13]. Dollberg et al. studied preterm infants to test the accuracy of digital thermometers that used predictive algorithms. In their study, the difference between measurements from the digital thermometer and a mercury thermometer was 0.14 °C [13]. Weiss and Richards tested both the "normal"—where the temperature measurement is based on the thermometer reaching thermal equilibrium—and predictive modes of an axillary thermometer to measure temperatures in neonates. The discrepancy between modes varied from 0.38 °C to −0.33 °C [10]. In another study, the difference between digital and mercury rectal thermometers in neonates was larger than 0.2 °C in 13% of the tested subjects [14]. The authors argued that a deviation of ±0.2 °C from the true body temperature may not be accurate enough to safely monitor sick or critical neonates. O'Brien et al., compared a predictive thermometer to an oral glass mercury thermometer to measure the oral temperature of 500 patients. The predictive thermometer accurately predicted fever only about 85% of the time; however, this means that it failed to detect 1 in every 7 fevers [15].

The referenced studies indicate that the predictive algorithms used in digital thermometers may not result in an accurate temperature measurement. The temperature measurements can be influenced by many environmental and anatomic factors, and a single predictive algorithm cannot account for all of the various conditions. Most importantly, none of the reviewed studies recorded data during the initial "transient" phase of the measurement, while the temperatures of the thermometers were increasing. These are the data that the digital thermometers apparently rely on as input for an exponential or other function to predict equilibrium temperature. Understanding the behavior of the thermometers during the initial temperature measurements is essential to understanding their ability to predict equilibrium temperatures.

The objective of this research was to conduct a clinical study on pediatric subjects to evaluate the accuracy of three brands of inexpensive, off-the-shelf digital electronic thermometers using an innovative approach. To understand the measurement accuracies of the off-the-shelf thermometer temperature predictions, we compared their predictions with those made simultaneously by a specially designed, calibrated, and fast-responding, computer-based, reference thermometer [16]. The reference thermometer provided improved data acquisition by measuring temperatures continuously, thus providing data during the entire transient phase of the measurement through to the equilibrium temperature. In this study, we directly compared continuously recorded clinical temperature reference measurements with those generated by the store brand thermometers. To the best of our knowledge, this was the first study using a fast responding and accurate reference thermometer to capture the transient behavior in body sites for body temperature measurements. The knowledge created by this study can be used to understand the thermometers' responses at different body sites. These results can form the basis for thermal modeling of the body and can be used to develop a clinically relevant test methodology to evaluate new thermometer designs resulting in more accurate, next-generation, low-cost digital electronic thermometers.

2 Methods

The procedures for the clinical study were approved by the Internal Review Boards at both UMBC (the University of Maryland, Baltimore County) and the U.S. FDA. No diagnoses or medical decisions were made based on the experimental measurements.

A total of 301 pediatric subjects, infants to 18 years old, participated in this study. The subjects were recruited during their visit to a local pediatrician's office (Box Hill Pediatrics, Abingdon, MD) for either well or sick visits. The subject's parent or guardian was given a consent form, and the procedure was explained. Once the appropriate consent form was signed by the parent or guardian, temperature measurements were taken in a normal examination room during the course of the patient's examination.

A reference thermistor bead temperature sensor ("reference thermometer") was specifically designed for this study. The reference probe works as a part of a commercially available temperature measurement system (T-View system, Alpha Technics, Irvine, CA). It has a nominal resolution of 0.001 °C. The time required for the reference thermometer to reach equilibrium at 37 °C when immersed into a 37 °C water bath (the "response time"), is less than 3 s due to its small size (<0.8 mm dia.). The reference thermometer was calibrated with a NIST traceable thermometer. The time constant of less than 5 s was determined using

a temperature controlled water bath [17]. Data are continuously recorded at an acquisition rate of 10 Hz. Therefore, the reference thermometer using a small thermistor bead temperature sensor is highly suitable for capturing transient behavior of thermal environment in biological systems when body temperature is measured.

The subjects' body temperatures were measured both under the subject's armpit (axillary site) and under their tongue (oral site). An exception to this was in the 0–2 yr patients, because it is not clinical practice to take oral temperature measurements in this age group. In addition, the 3–6 yr sick patients generally did not feel well enough for the oral measurements. The tested store brand thermometer and the reference thermometer were placed together into a disposable oral thermometer sheath (TIDI Products, Neenah, WI) and placed simultaneously under the subject's arm to record temperatures for 100–120 s. Next, the thermometers were placed together into a new disposable sheath, and simultaneously placed under the subject's tongue to record temperatures for another two minutes. All temperature data for both the reference and the store brand thermometers, and subject's age and gender, were recorded. Temperatures were measured and recorded in Fahrenheit, as that is the clinical standard for temperature measurements in the U.S. Temperatures were converted to °C, and both values are reported (e.g., 100 °F (37.8 °C)).

Three brands of store brand digital thermometers were used (labeled here as A, B, and C). All three store brand thermometers were tested on similar numbers of patients (Brand A: 102; Brand B: 105; Brand C: 94). Up to 28 thermometer probes within each brand were evaluated to test consistency within a brand.

The subjects were nominally divided into four age groups (0–2, 3–6, 7–12, and 13–18 yr) for initial data analysis. Within each age group, subjects were categorized as either sick or healthy, depending on the parents' or guardians' self-reported reason for their visit. The pediatric clinic's staff did not routinely measure the body temperature of the patient; clinical temperature was measured only when the subject checked in as sick. The standard body temperature measurement device used in the pediatrician's office was a temporal temperature scanner, the accuracy of which was unknown to us [18,19]. No mercury thermometer was used in the clinic. During the time frame of the clinical study, the number of individuals in each of the age groups depended on the subjects visiting the clinic, and therefore varied from week to week. We found that the subjects in the 0–2, and some in the 3–6, year old age group did not cooperate, and refused to allow thermometers into their mouths; therefore, no oral temperature data were collected for those subjects. In addition, as suggested by Latman [20], the data were considered invalid when the sensor or thermometer was not properly placed in the measuring site and/or the patient did not cooperate.

Temperatures measured for each age group, measurement site, as well as temperature differences between specific groups, are presented as mean ± standard deviation ($d \pm$ SD). We assumed that the temperature distribution was Gaussian; therefore, 95% of the samples lie between d-1.96 SD and $d + 1.96$ SD. If the measurements are different between two thermometers or two measurement sites, the range between d-1.96 SD and $d + 1.96$ SD is defined as the limits of agreement (LOA). The LOA give the upper and lower range with 95% confidence [21,22]. The LOA provide a robust method to quantify the differences between a measurement method and a standard, or reference, measurement method. Limits of agreement are presented as [d-1.96 SD, $d + 1.96$ SD] in this report (e.g., [2.24,−1.25 °F], ([1.24, −0.69 °C])).

The continuously recorded temperature versus time data (the temperature rise curve), measured by the reference thermometer, also provided detailed information on how the temperature profile at the measuring site recovered after, for example, opening and closing the mouth or unfolding and closing the arm. The transient temperature data were evaluated post hoc to determine whether an exponential function can describe the temperature rise curve using a single characteristic time constant. The maximum temperature (T_{max}) was defined as the highest temperature reached during the 2 min of recording. Two characteristic temperatures were determined for each reading: $t_{0.5}$ was defined as the time when the temperature reached half of the maximum temperature, measured from the baseline (room temperature, T_{air}), and $t_{0.63}$ is the time when the temperature reached 63% of the maximum temperature. In a dynamic process described by an exponential function, $t_{0.63}$ is usually considered to be the characteristic time constant of the transient process. The $t_{0.63}$ time constants were used to compare the temperature rise curves in different locations, as well as between age groups and patient condition (healthy or sick).

A post hoc analysis of the time constant $t_{0.63}$ was performed for a random subset of the data to test whether the reference thermometer's transient response was different for different age groups. To check the homogeneity of variance we used the Bartlett test. To check for differences between means of different age groups, a one-way analysis of variance (ANOVA) was performed. If the ANOVA showed that there were significant differences between means, then a subsequent test for comparing means between multiple groups was performed (Tukey test). All statistical analyses were performed using R software (The R Project). Differences were considered significant for $p < 0.05$.

3 Results and Discussion

3.1 Experimental Groups and Sample Sizes. Table 1 shows the experimental groups and number of participants in each group. The sample size was typically smaller in the oral groups ($n = 2$–17), since the thermometer sometimes made the patients uncomfortable, especially children of younger ages. Of the 301 subjects that participated, 212 were healthy (70%) and 89 were sick (30%).

3.2 Temperature Transients Measured by the Reference Thermometer. We recorded the temperature as a function of time for 2 min after the reference thermometer was placed in the measuring site. The data for the healthy 7–12 year olds varied from one patient to another in the initial 0–20 s of measurement, with a variation of 5–10 °F (2.8–5.6 °C) (Fig. 1). This may be largely due to how fast the store brand and reference thermometers are inserted, for how long the arm or mouth is unfolded or open, and whether the patient remained in the same position for the two minute measurement. However, the variations decreased as time passed. Extrapolating from the transient curves, it appears that the temperatures would have continued to increase slightly after the approximately 2 min of our measurements in some cases.

The reference thermometer measurements provided a novel basis to compare the results of the different store brand thermometers, which were used on different patients at different times. In data that can be represented by an exponential function with a single time constant, τ, the value is 63% of the maximum value at $t = \tau$, and 99% of the maximum value at $t = 4.6\tau$. Therefore, if an exponential function with a single characteristic time constant describes the transient behaviors shown in Fig. 1, the temperature sensed by the reference thermometer should reach 99% of its maximum value after $4.6t_{0.63}$. In Fig. 1, after a time equivalent to $4.6t_{0.63}$ in each group (approximately 20–34 s—represented by the vertical line), the temperatures had generally not reached 99% of their maximums, implying that the temperature transients may not be well described by an exponential function with only a single characteristic time constant. The use of a single time constant is the result of applying the lumped capacitance method, which assumes that a solid with a uniform temperature is immersed in a fluid at a different temperature [23]. Because a uniform temperature in the solid is assumed, there can be no contribution of any temperature gradient in the solid. A temperature gradient at the clinical measurement site is a certainty. For example, a continuous increase in the blood perfusion rate at the measuring site

Table 1 The experimental groups and number of participants in each group

	Age (years old)	0–2		3–6		7–12		13–18	
Thermometer[a]	Healthy/Sick	Healthy	Sick	Healthy	Sick	Healthy	Sick	Healthy	Sick
A	Oral	0	0	6	2	17	11	17	3
	Axillary	10	2	18	7	25	13	24	4
B	Oral	0	0	5	5	15	11	20	6
	Axillary	8	3	16	13	19	14	24	7
C	Oral	0	0	6	0	18	2	10	7
	Axillary	3	4	27	10	23	3	15	9

[a]Note that each store brand thermometer measurement was matched with a reference thermometer measurement. Note also that typically, in the clinical situation, oral temperature measurements are not taken for children in the 0–2 yr age group. In addition, the sick children in the 3–6 yr age group were reluctant to have their temperature taken orally.

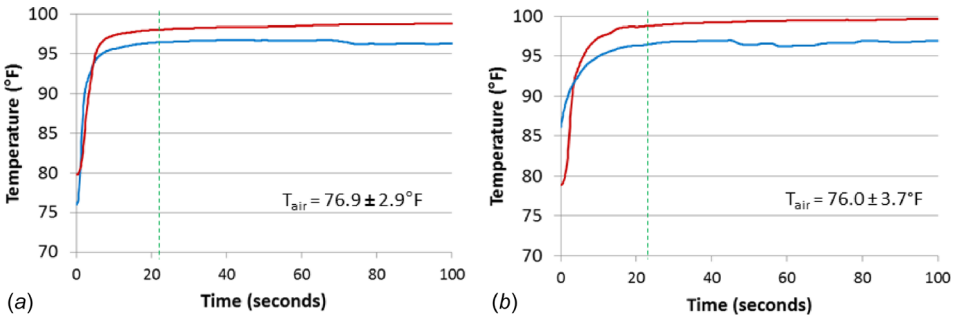


Fig. 1 Examples of temperature transient data recorded by the reference thermometer at the oral site in healthy patients in the 7–12 year old groups. The two data sets on each graph represent the transient behavior for the highest and lowest measured temperatures at 100 s. There was essentially a continuum of different transient responses and values of *T* at 100 s (shown in the Appendix). The vertical dashed line in each panel represents the average values of the 4.6$t_{0.63}$. These curves represent the transient response upon which the store brand thermometers base their predictions. Graph (*a*) shows two similar transient curves rising in parallel. The curves start at different temperatures based on environmental and patient conditions. Graph (*b*) shows the transient behavior of the two curves crossing, where the lower final temperature curve actually started at a higher initial temperature.

(especially at the axillary site, when the skin is exposed to air) may play a significant role in the slow temperature increases observed as the thermometer reading reached body temperature over a long measurement duration. Therefore, two thermal mechanisms—(1) opening and closing the arm, and (2) local blood perfusion rate increase—likely contribute to, and have an effect on, the transient thermal temperature profiles, and therefore the transient behavior cannot be modeled by an exponential function with a single time constant. We previously proposed a model that included two exponential terms, each with a unique time constant that accounted for the different bioheat transfer mechanisms and therefore better approximated the transient temperature rise [24].

The relationship between the patient's age and the reference thermometer's time constant $t_{0.63}$, for a random subgroup of subjects, is presented in Fig. 2. The average time constant was larger when the patients were younger: $t_{0.63} = 5.67$ s in the 0–2 yr group, in contrast to $t_{0.63} = 4.14$ s in the 13–18 yr group. This is expected because, in general, the older the child, the larger the body and arm. A bigger body size (or a bigger arm) should have a larger thermal capacity, and therefore be less influenced by the cooler ambient environment in the room during a transient process. However, the one way ANOVA and Bartlett's test showed that there were no statistically significant differences in the mean values or the variances between any of the age groups. This indicates that the reference thermometer's transient response was consistent across all age groups. Thus, by having the reference thermometer measuring temperature simultaneously with the test thermometer for each subject, the difference in the temperature measured between the two thermometer readings normalizes patient to patient variation, and is a robust measure of variation in the test thermometers.

3.3 Average Temperatures Measured by the Reference and Store Brand Thermometers. Figure 3 gives the temperature values recorded at the oral site by the reference and the Brand A thermometers in the three age groups. After two minutes, the temperature at the oral site ranged from 97.6 °F (36.4 °C) to 98.2 °F (36.8 °C). The temperature was approximately 0.5 °F (0.28 °C) lower for the measurements recorded after 1 min. After approximately 10 s, when a typical store brand thermometer stops recording, the reference temperature was approximately 3 °F (1.7 °C) lower than that measured at 2 min. A large standard deviation (up

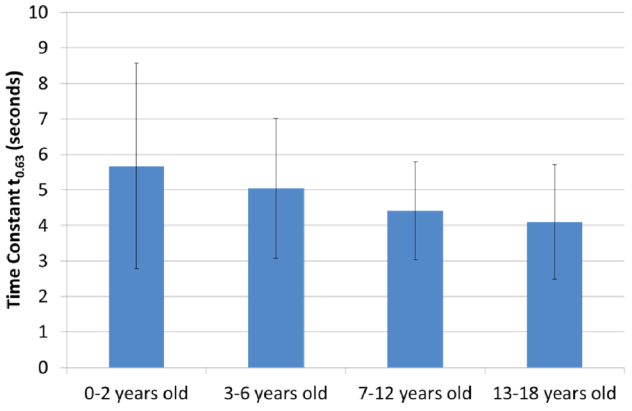


Fig. 2 Time constant $t_{0.63}$ at the axillary site in different age groups of healthy patients. Error bars represent standard deviation.

to ±3.2 °F (1.8 °C)) was found after 10 s, but gradually decreased, indicating that the most significant temperature variation among patients occurred during the first few seconds. The temperature values measured within the initial 10 s are the data used by store brand thermometers for predicting the body temperature. At 2 min, the standard deviation of temperature measured by the reference thermometer is much smaller (Fig. 3); however, the large variations present in the initial 10 s are not reflected in the temperature predictions by the store brand thermometer. The average of the temperature measurements predicted by the Brand A thermometers is not statistically significantly different from the average of the temperature measurements from the reference thermometer, for any of the age groups. The average temperature measurements from Brands B and C thermometers followed a similar pattern and were also not significantly different from the average temperature measurements of the reference thermometer (Appendix).

The recorded temperature trends at the axillary site are presented in Fig. 4. The temperature rises more quickly in the older age groups, consistent with the trend of the calculated $t_{0.63}$ values in Fig. 2. After two minutes, the average temperatures recorded by the reference thermometer (97.1 °F to 97.5 °F (36.2 °C to 36.4 °C)) are lower than those recorded at the oral site (97.6 to 98.2 °F (36.4 to 36.8 °C), Fig. 3). The average temperatures predicted by the store brand thermometers are not statistically significantly different from those measured by the reference thermometer after 2 min (the same was true for Brands B and C, see Appendix). Due to the slower temperature rises at the axillary site, the temperatures predicted by the store brand thermometer (97.1–97.4 °F (36.2–36.3 °C)) were lower than those measured at the oral site (97.8–98.1 °F (36.6–36.7 °C)). This seems reasonable because the axillary site underestimates the body core temperature more than the oral site. Large temperature variations also existed at the axillary site after 10 s (up to ±2.2 °F (1.2 °C)). Again, the large standard deviation at 10 s is not reflected by the store brand thermometer measurements, since their predictions have a much smaller standard deviation (±0.6 °F (0.33 °C)).

The temperature measurements shown in Figs. 3 and 4 are those of healthy patients. The recorded temperatures for the sick patients in two age groups are illustrated in Fig. 5. Overall, the standard deviations of the reference thermometer data in the sick groups (0.7–1.4 °F (0.39–0.78 °C)) are larger than those in the healthy groups (0.7–1.0 °F (0.39–0.56 °C)); however, again the average temperature measurements shown for the Brand A thermometers (and also for Brands B and C, see Appendix) were not significantly different from the average temperature measurements of the reference thermometer.

3.4 Differences in Temperatures Measured by the Reference and Store Brand Thermometers. For the three age groups where patients had both oral and axillary temperature measurements, Table 2 gives the average temperature differences between the oral and axillary temperatures measured by the reference thermometer. The average differences from the reference thermometer between the two sites are all negative, suggesting that the measured axillary temperature is lower than the oral temperature of the same patient.

Figure 6 presents temperature differences between the store brand thermometer in the predictive mode and the matched reference thermometer after two minutes at the axillary sites for each trial in the healthy patients. Positive differences mean that the store brand thermometer measurement was higher than that from the reference thermometer, and negative differences mean that the store brand measurement was lower. For both the healthy patients

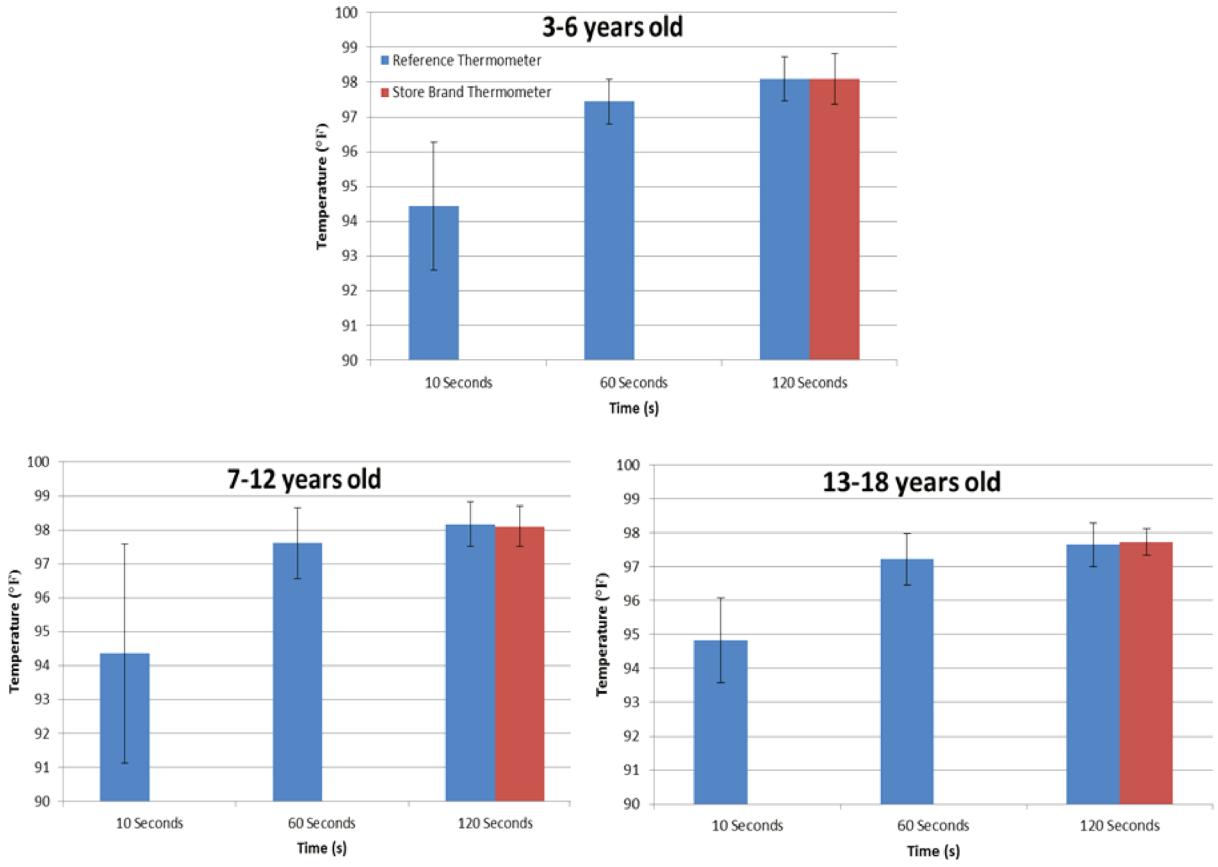


Fig. 3 Average temperatures at various time instants measured by the reference thermometer and the predicted temperature by the store brand A thermometers. The data are presented for results from healthy patients in three age groups at the oral site. The average temperature measurements at 120 s were not statistically significantly different.

(Fig. 6(*a*)) and sick patients (Fig. 6(*b*)), the data scatter about the *x*-axis with an almost equal number of positive and negative differences. This suggests that in approximately half of the cases, the store brand thermometer overestimated, and in half the cases it underestimated, the temperature.

For the healthy patients, the LOA's (presented as [d-1.96 SD, d + 1.96 SD]) for Brand A were, in order of range from greatest to least: 0–2 yr [2.24, −1.25 °F] ([1.24, −1.25 °C), 7–12 yr [1.33, −1.32 °F] ([0.74, −0.73 °C]), 3–6 yr [1.18, −1.28 °F] ([0.66, −0.71 °C]), and 13–18 yr [0.64, −0.92 °F] ([0.36, −0.51 °C). The Brand A thermometers were the most inaccurate in the 0–2 age group, with an average difference of ±1.9 °F(1.1 °C), and an average difference of at least ±1.2 °F (0.67 °C) for the other three age groups.

For Brand B, the LOA in order of range from greatest to least: 3–6 yr [1.97, −2.16 °F] ([1.1, −1.2 °C), 7–12 yr [1.00, −1.88 °F] ([0.56, −1.04 °C]), and 13–18 yr [0.82, −1.68 °F] ([0.46, −0.93 °C]). It was evident that the Brand B thermometers underestimated the temperature at the axillary site most of the time. In contrast, the mean values of the temperature differences of Brand C varied from 0.95 °F (0.53 °C) (13–18 age group) to 1.38 °F (0.77 °C) (3–6 age group). The LOA's for Brand C were, in order of range from greatest to least: 7–12 yr [3.71, −1.43 °F] ([2.06, −0.79 °C), 3–6 yr [3.29, −0.46 °F] ([1.83, −0.26 °C]), 13–18 yr [2.17, −0.25 °F] ([1.21, −0.14 °C]), and 0–2 yr [1.87, −0.16 °F] ([1.04, −0.089 °C]). The Brand C thermometers generally overestimated the reference axillary temperature, by up to 3.7 °F (2.06 °C) in one instance.

For sick patients Brand A thermometers again deviated from the reference temperature on both the positive and negative sides, with an average difference equal to 1.1 °F (0.61 °C). The LOA's for Brand A were, in order of range from greatest to least: 7–12 yr [1.66, −1.91 °F] ([0.92, −1.06 °C]), 0–2 yr [0.92, −1.15 °F] ([0.51, −0.64 °C]), 3–6 yr [1.24, −0.59 °F] ([0.69, −0.33 °C]), and 13–18 yr [0.18, −0.52 °F] ([0.1, −0.29 °C]). Brand B thermometers underestimated, more than overestimated, the reference axillary temperature. The LOA's for Brand B were, in order of range from greatest to least: 3–6 yr [1.04, −1.42 °F] ([0.58, −0.79 °C]), 7–12 yr [0.95, −0.95 °F] ([0.53, −0.53 °C]), 0–2 yr [0.38, −1.3 °F] ([0.21, −0.72 °C]), and 13–18 yr [0.47, −0.68 °F] ([0.26, −0.38 °C]). The Brand C thermometers overestimated the temperatures 92% of the time, by up to 3.5 °F in one instance. The LOA's for Brand C were, in order of range from greatest to least: 7–12 yr [4.38, −0.72 °F] ([2.43, −0.40 °C]), 0–2 yr [3.99, 0.72 °F] ([2.22, −0.4 °C]), 3–6 yr [3.46, −0.75 °F] ([1.92, −0.42 °C]), and 13–18 yr [1.62, −0.88 °F] ([0.9, −0.49 °C]). The data show that while the average temperatures measured by the reference and store brand thermometers were similar (Figs. 3–5), the average temperature measurements do not reflect the accuracy of the temperature measurements. The differences between the reference and store brand thermometers were inconsistent, and ranged between −1.4 °F (−0.78 °C) and +3.5 °F (1.94 °C). The relevant working range of a clinical thermometer is approximately 95–105 °F (35–40.6 °C), not ~ 0–100 °F (0–37.8 °C); therefore, differences of ±1 or 2 °F (0.56 or 1.12 °C) represent discrepancies from 10 to 20%.

The data and trends from the oral sites were similar, for both healthy and sick patients. We obtained fewer data points, especially in the younger patients, who would often refuse to place both thermometers in their mouths. The complete data can be found in the Appendix.

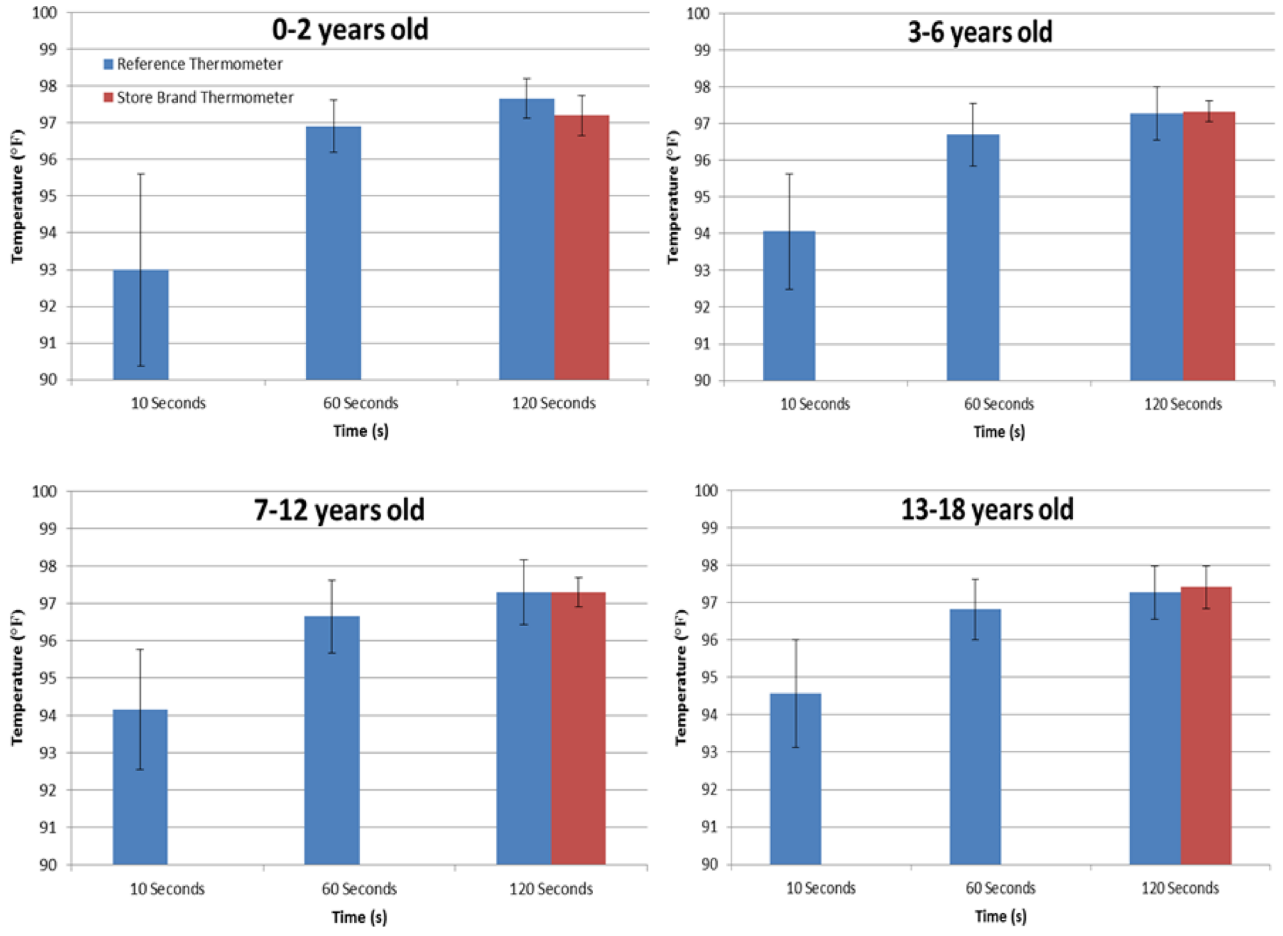


Fig. 4 Average temperatures at various time instants measured by the reference thermometer and the predicted temperature by the store Brand A thermometers. The data are presented for results from healthy patients in four age groups at the axillary site. The average temperature measurements at 120 s were not statistically significantly different.

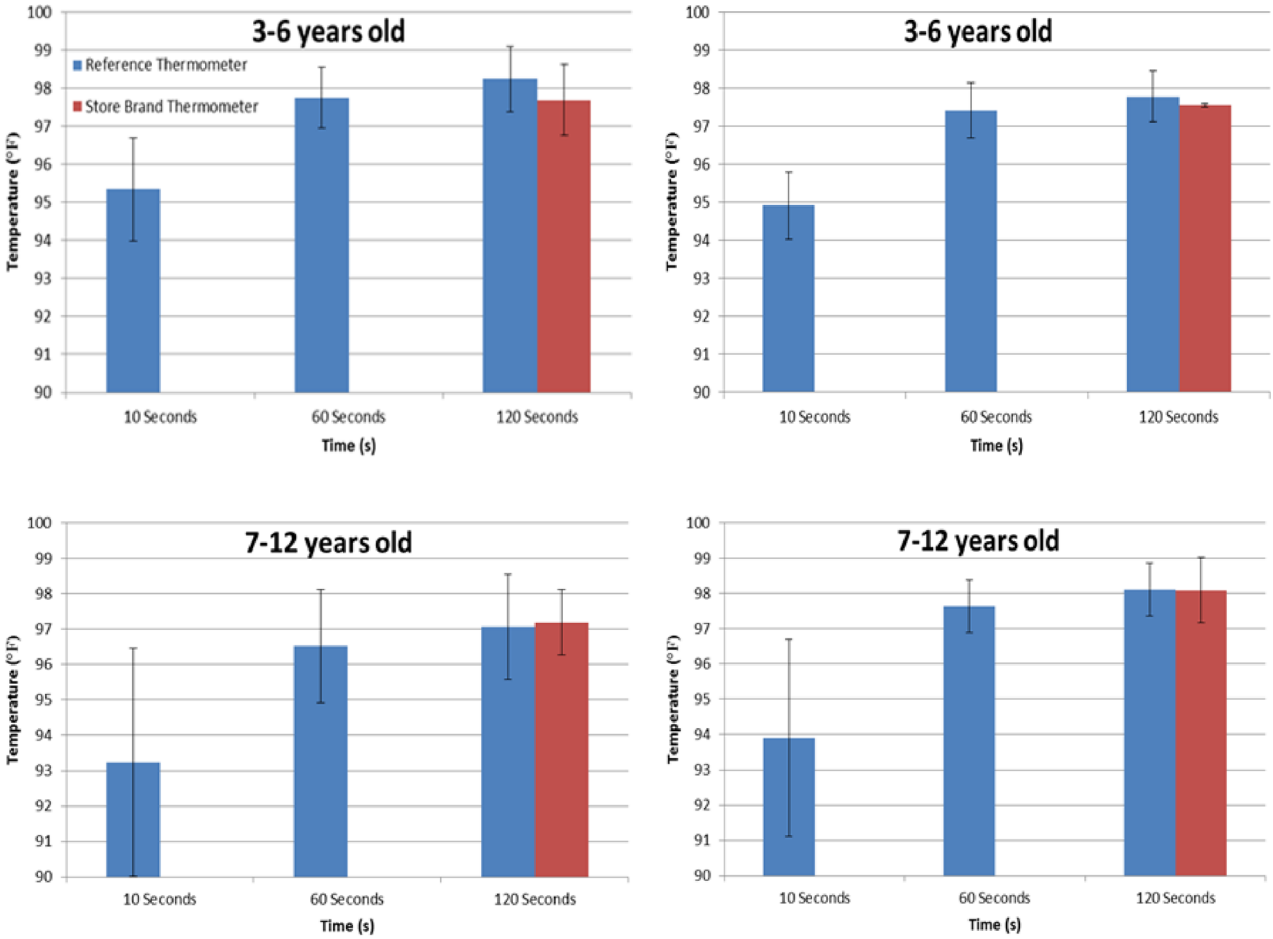


Fig. 5 Average temperatures at various time instants measured by the reference thermometer and the predicted temperature by the store Brand A thermometers. The data are presented for results from sick patients in two age groups at the axillary (left graphs) and oral (right graphs) sites. The average temperature measurements at 120 s were not statistically significantly different.

The store brand thermometers in predictive mode necessarily use an extrapolation algorithm to predict the core body temperature after only a few seconds of measurement. The extrapolation algorithm(s) used in the store brand thermometers are unknown. For example, they may use a linear extrapolation, simply add a constant to the measured temperature after 5 s, or use a more complex function (e.g., exponential). Our reference thermometer data show that the initial rise in temperature is highly inconsistent (Fig. 1) and is not a good predictor of the ultimate, equilibrated temperature. Our data suggest that a more complete description of the thermal processes should be incorporated into any predictive algorithm. In a previous study, we suggested an equation based on heat transfer theories [24]. This equation is a combination of two exponential functions with different time constants that better captures the transient behavior of a thermometer during temperature measurement:

$$\frac{T(t)-T_0}{T_{ss}-T_0} = A[1-\exp(-t/\tau_1)] + (1-A)[1-\exp(-t/\tau_2)] \quad (1)$$

where T_0 is the ambient room temperature, T_{ss} is the local steady-state temperature at the measurement site t is time, and τ_1 and τ_2 are different time constants. The left side of Eq. (1) defines the dimensionless temperature and varies between zero and one. Each of the two exponential functions on the right side of Eq. (1) are defined with the different time constants, τ_1 and τ_2 (where $\tau_1 < \tau_2$). A is a constant that defines the relative contribution of the first exponential term, which describes the transient behavior of the temperature rise, and $1\text{-}A$ defines the relative contribution of the second term, which describes the longer-term behavior as the temperature approaches steady-state. We believe that incorporating such an algorithm could improve the accuracy of fast read thermometers in the predictive mode.

Table 2 Difference between the axillary and oral ($T_{axillary} - T_{oral}$) temperatures of the same subject in three age groups measured by either the reference thermometer or a store brand thermometer

	3–6 years old	7–12 years old	13–18 years old
Reference thermometer	−0.773 ± 0.997 (n = 16)	−0.786 ± 0.902 (n = 49)	−0.460 ± 0.727 (n = 46)
Brand A	−0.833 ± 1.05 °F (n = 6)	−0.875 ± 0.543 °F (n = 17)	−0.45 ± 0.581 °F (n = 16)
Brand B	−0.54 ± 0.38 °F (n = 5)	−0.471 ± 0.658 °F (n = 14)	−0.14 ± 0.707 °F (n = 20)
Brand C	0.04 ± 0.926 °F (n = 5)	−0.759 ± 1.43 °F (n = 17)	−0.540 ± 0.524 °F (n = 10)

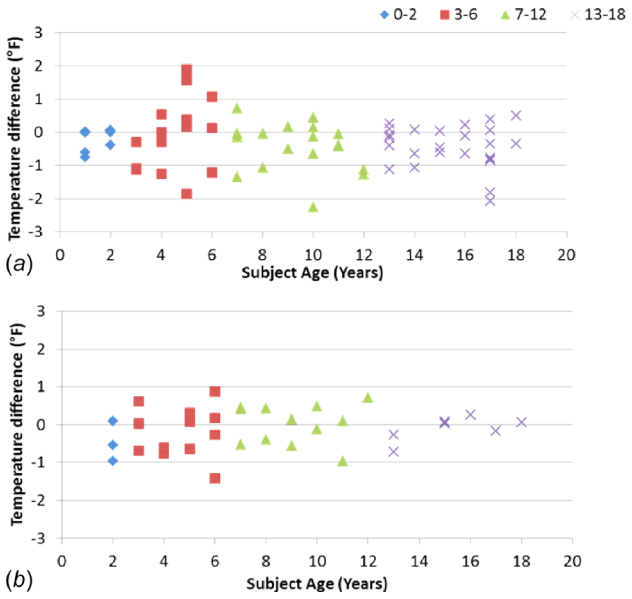


Fig. 6 An example of the ranges of temperature differences between the reference thermometer at two minutes and store brand thermometer (B) at the axillary sites for the healthy patients (*a*), and sick patients (*b*). Positive temperature differences mean that the temperature sensed by the store brand thermometer is larger than that by the reference thermometer. Complete data at both the axillary and oral sites are given in the Appendix.

3.5 Variations Within Each Store Brand. The large variations between the store brand and the reference thermometers may be due to large discrepancies between individual thermometers, or even defective thermometers. To control for the possibility of defective thermometers, we tested how individual store brand thermometers deviated from the reference thermometer results. Figure 7 presents the data from Fig. 6 replotted to show the measurement differences between individual store brand thermometers (Brand B) and the matching reference thermometer measurement. More than 18 thermometers were tested for each store brand at the axillary site. Some thermometers were used in more than nine patients, while others were used only once or twice. It is evident from the data that a single thermometer could both under- and over-estimate the temperatures at the axillary site. The deviations were distributed randomly. Analogous data for sick subjects at the axillary site, and both healthy and sick subjects at the oral site, are given in the Appendix.

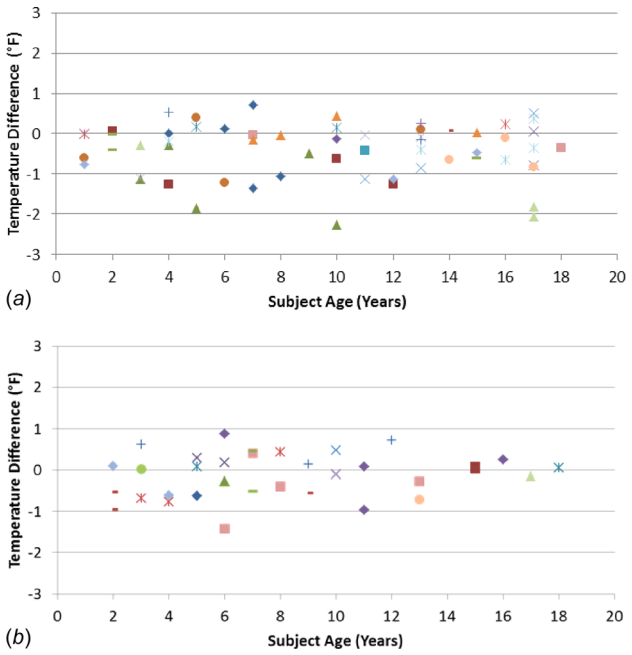


Fig. 7 Temperature difference data from each store brand thermometer (probe) showing the dispersion of temperature differences between the reference thermometer at two minutes and a store brand thermometer (B) at the axillary sites for the healthy (*a*) and sick (*b*) patients. Each individual thermometer (probe) is represented by a unique symbol. Note the inconsistency in individual thermometer results. Complete data at both the axillary and oral sites are given in Appendix.

4 Conclusions

To the best of our knowledge, this study was the first to investigate the temperatures measured by off-the-shelf thermometers compared to a reference thermometer—a thermistor bead sensor—in a clinical setting with a large number of participants. The reference thermometer and the store brand thermometer measured temperatures simultaneously to evaluate the predicted values of the store brand, and to account for patient-to-patient variations. The data generated, therefore, represent unique information to enable the design of more accurate, next generation commercially available thermometers.

Compared to the reference thermometer measurements after 120 s (two minutes), the store brand thermometers routinely deviated from the reference temperature, and those deviations were not consistent. The Brand C thermometers had the greatest deviations of up to 3.7 °F (2.1 °C), while the Brand A thermometers had the lowest deviations; however, they still deviated by up to 1.9 °F (1.1 °C). In this study, the tested store brand thermometers had inaccuracies greater than the indicated ±0.2 °F (0.11 °C) compared to the reference thermometer after 2 min. Our recorded transient data showed that there was a wide variation in the transient temperature profiles. Significant deviations from the maximum temperature after $t = 4.6t_{0.63}$ illustrated that the transient temperature profiles may not be represented by an exponential function with a single characteristic time constant, suggesting that more than one thermal mechanism is involved during the transient process. This insight is critical to the design of improved clinical thermometers. The store brand thermometers state that they predict body temperature based on transient temperature measurements over the first 5–10 s, implying that they use an embedded algorithm or correction to calculate the body temperature. The accuracy of the embedded algorithms was not verified directly by our study, since the predicted body temperatures do not capture the large variations observed over the initial 10 s of the measurements. A thermometer with an unpredictable error of plus or minus several degrees Fahrenheit may result in a false positive or negative diagnosis of fever in children. The results of this study, specifically the method of recording real-time transient temperature data simultaneously with either existing thermometers or with new designs, may provide a more robust method to understand critical design parameters to improve measurement accuracy in next generation thermometers. The next step for thermometer manufacturers, the research community, and the FDA is to develop a more appropriate standardized test, i.e., a new ASTM or ISO standard to ensure the accuracy and consistency of these thermometers.

Acknowledgment

This study was supported by the U.S. FDA. We wish to thank the Physicians and Staff of Box Hill Pediatrics for allowing us to perform the clinical study in their offices, and for their help in interpreting medical information. We would also like to thank Dr. Nagaraj K. Neerchal of the Department of Mathematics and Statistics at the University of Maryland, Baltimore County, for his help in performing and interpreting the statistical analyses.

Appendix

The appendix contains data not presented in main text. Figure numbers correspond to the figure numbers of the analogous data presented in the main text.

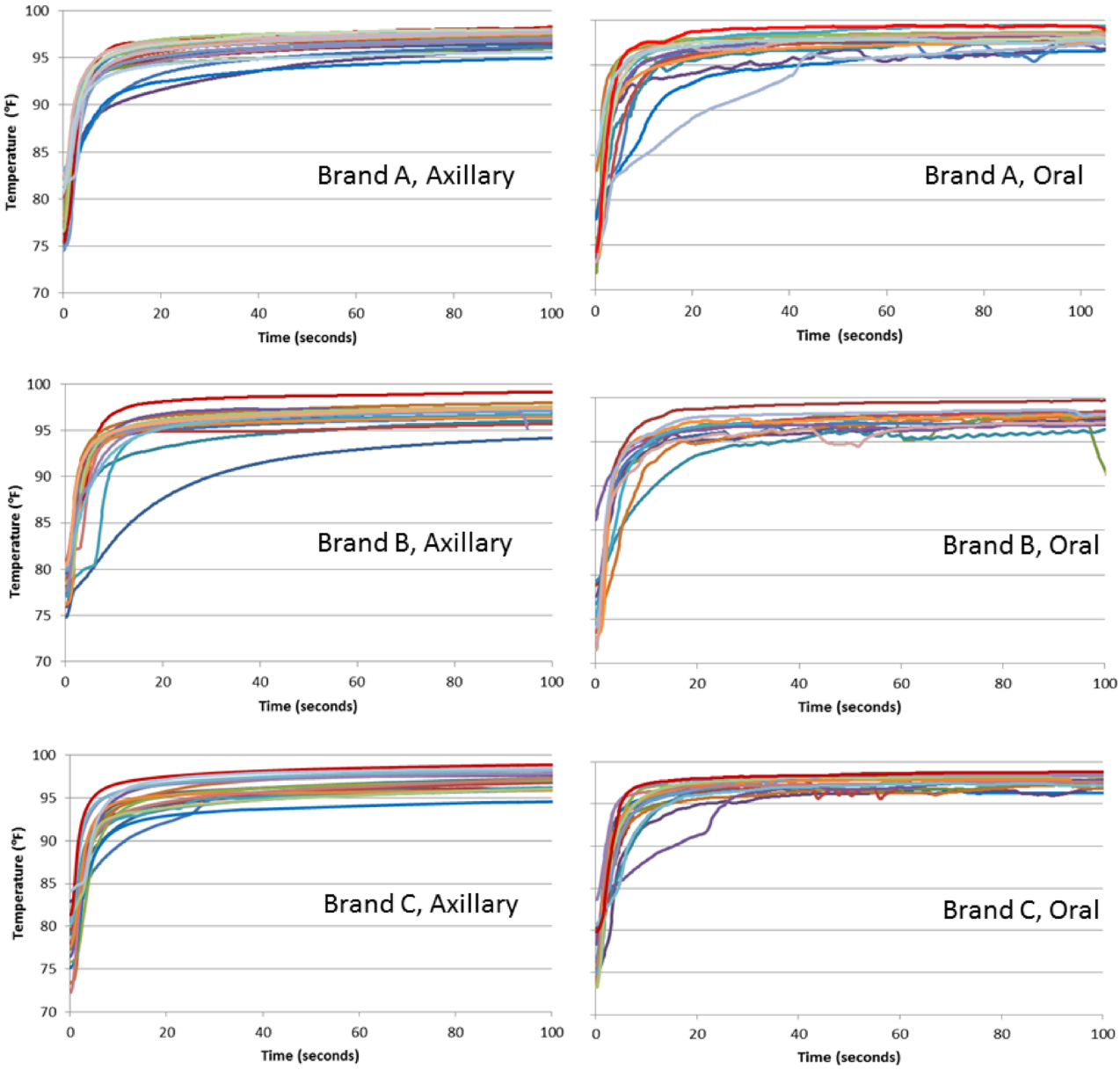


Fig. 8 Examples of temperature transient data recorded by the reference thermometer in healthy patients in the 7–12 year old groups. These plots show essentially a continuum of different transient responses and values of *T* at 100 s. The vertical dashed line in each panel represents the average values of the $4.6t_{0.63}$. These curves represent the transient response upon which the store brand thermometers base their predictions.

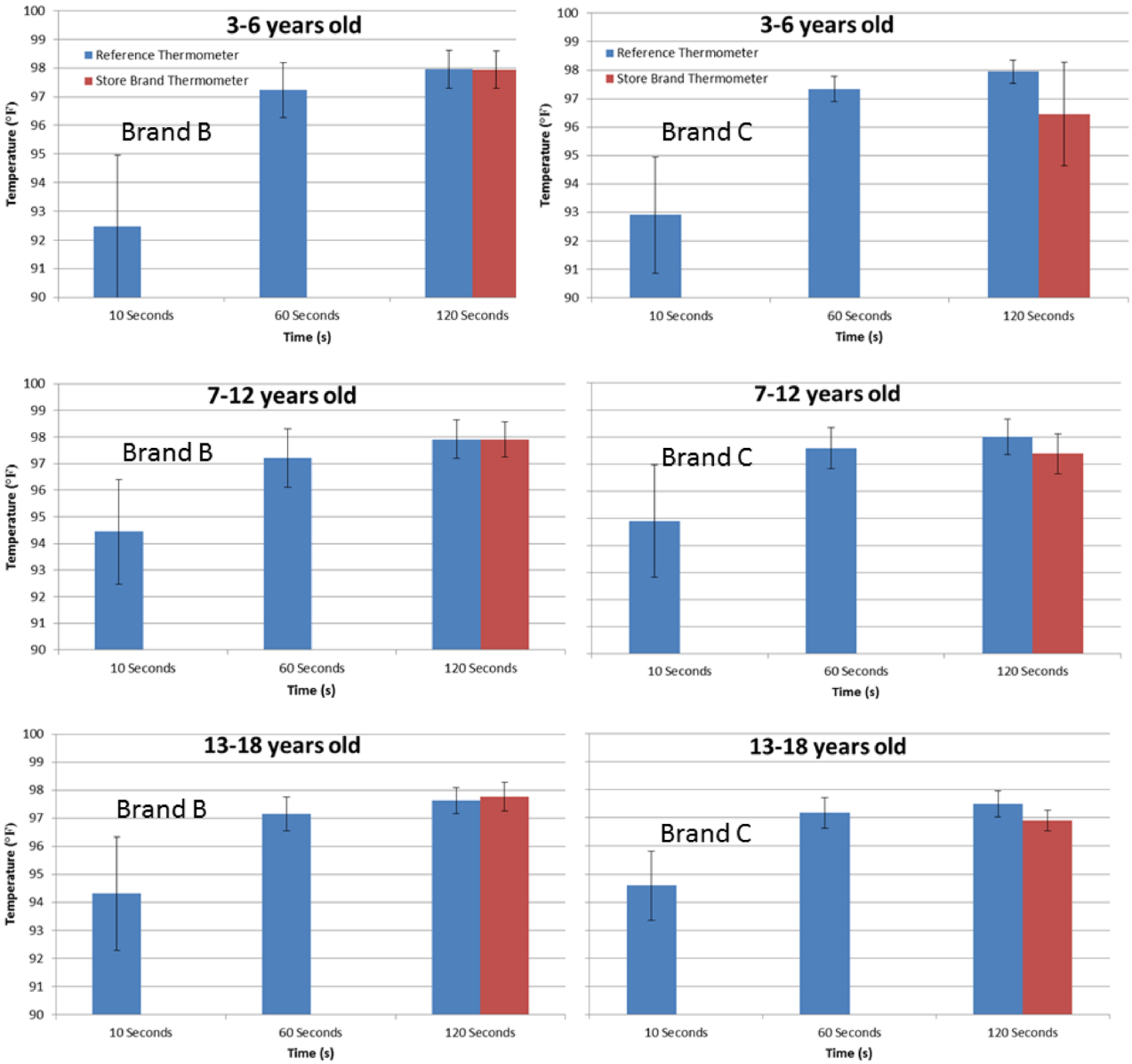


Fig. 9 Average temperatures at 10, 60, and 120 s measured by the reference thermometer and temperature predicted by the store Brand B and C thermometers in the fast predictive mode. The data are presented for results from healthy patients in three age groups at the oral site (it was not possible to collect oral measurements in the 0–2 yr age group). The average temperature measurements at 120 s were not statistically significantly different.

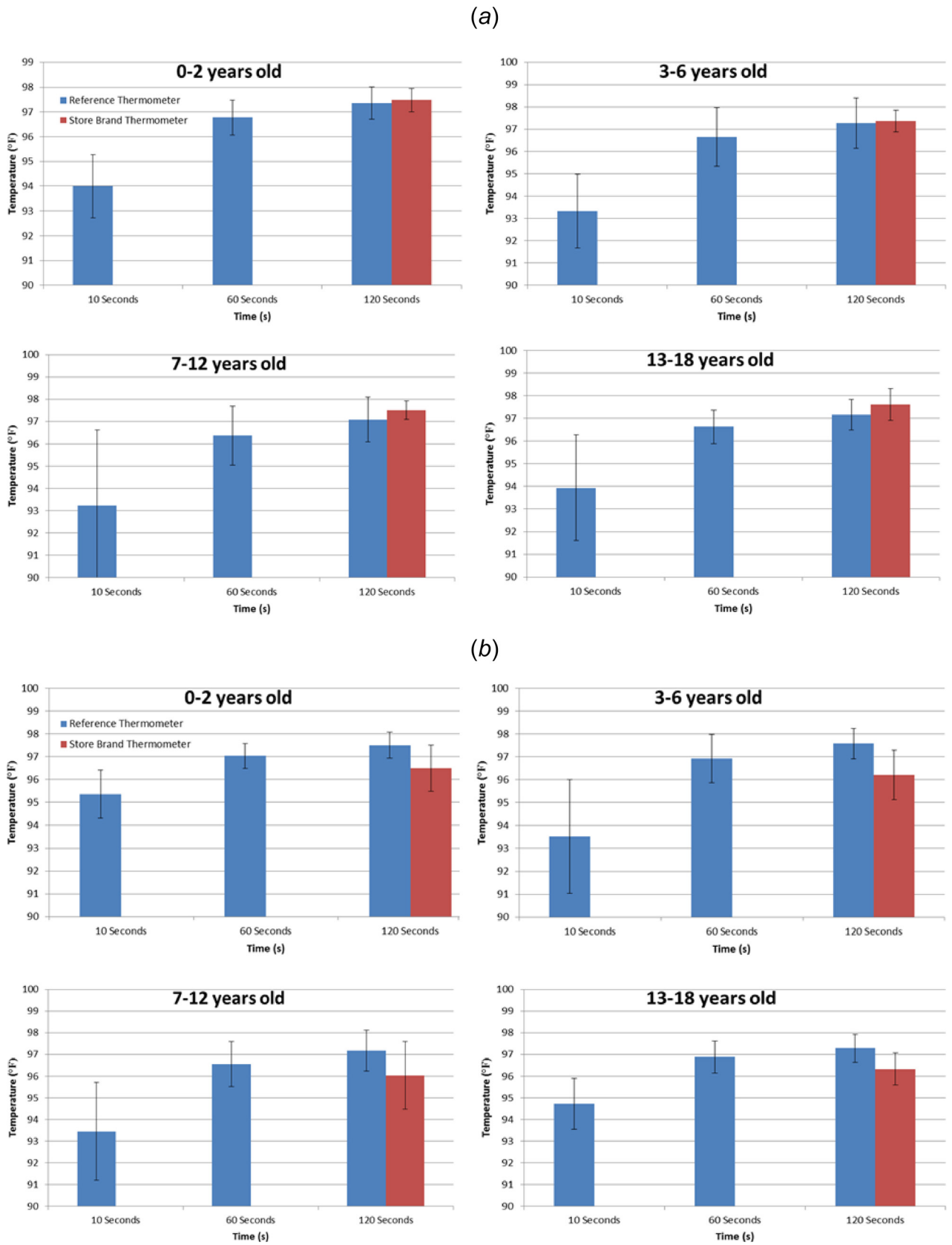


Fig. 10 Average temperatures at 10, 60, and 120 s measured by the reference thermometer and the predicted temperature by the store Brand B and C thermometers. The data are presented for results from healthy patients in four age groups at the axillary site. The average temperature measurements at 120 s were not statistically significantly different: (*a*) Brand B and (*b*) Brand C.

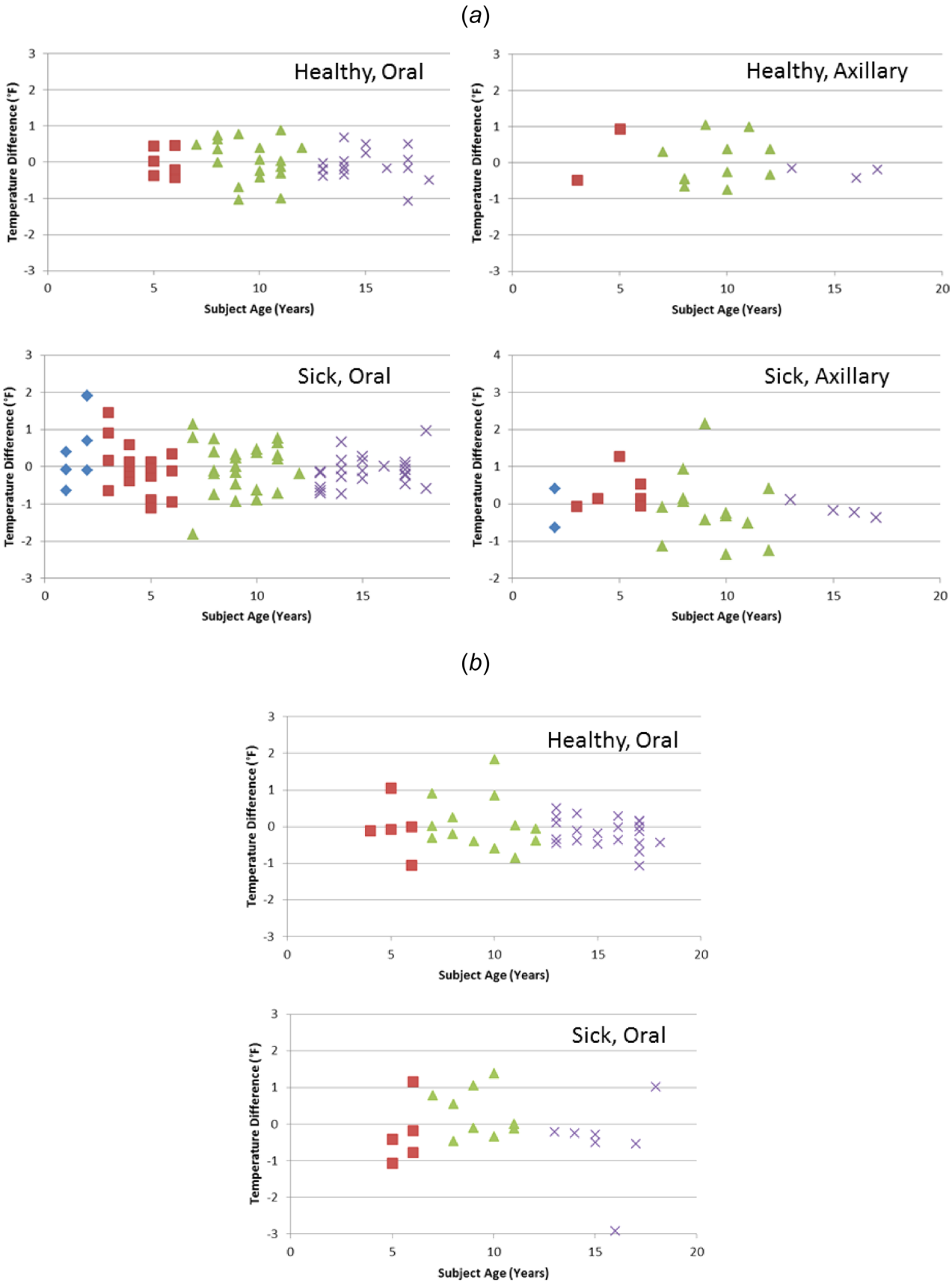


Fig. 11 Temperature differences between the reference thermometer at two minutes and the store brand thermometers at the oral site for the healthy and sick patients. Positive temperature differences mean that the temperature sensed by the store brand thermometer is larger than that by the reference thermometer: (*a*) Brand A, (*b*) Brand B, and (*c*) Brand C.

(*c*)

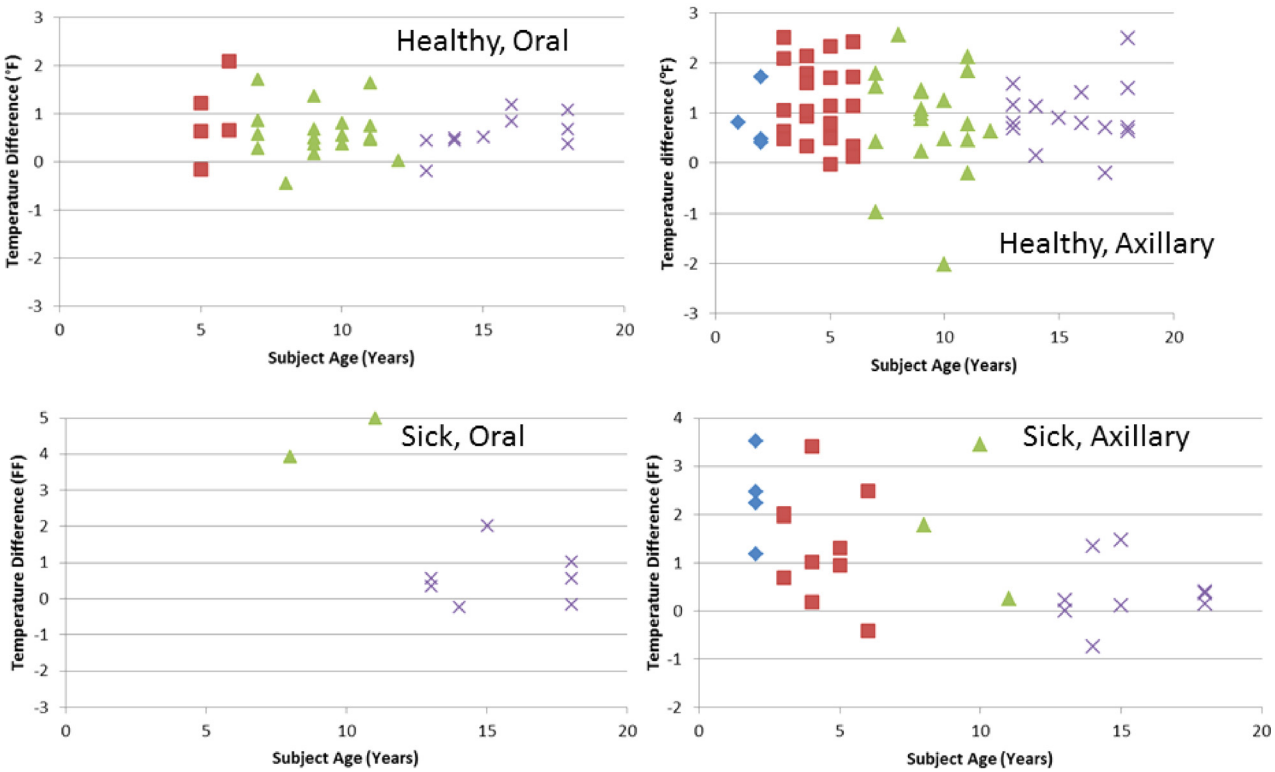
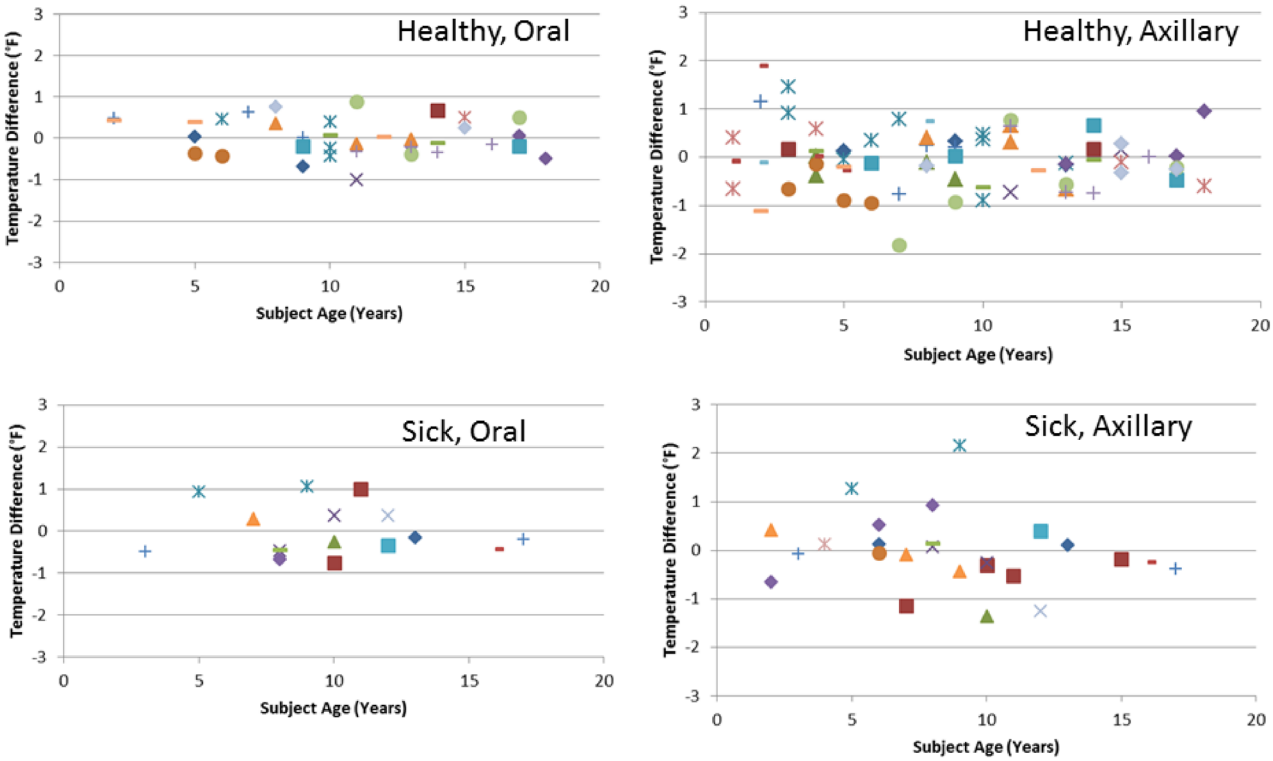


Fig. 11 Continued

(*a*)



(*b*)

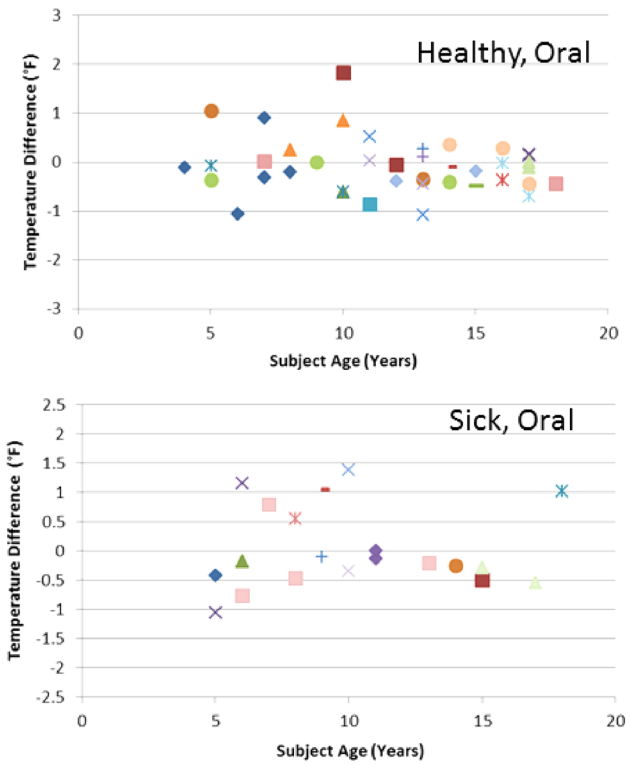


Fig. 12 Data from each store brand thermometer (probe) showing the temperature differences between the reference thermometer at two minutes and the store brand thermometers at the oral (healthy and sick) and axillary (sick) sites. Each individual thermometer (probe) is represented by a unique symbol. Note the inconsistency in individual thermometer results: (*a*) Brand A, (*b*) Brand B, and (*c*) Brand C.

(c)

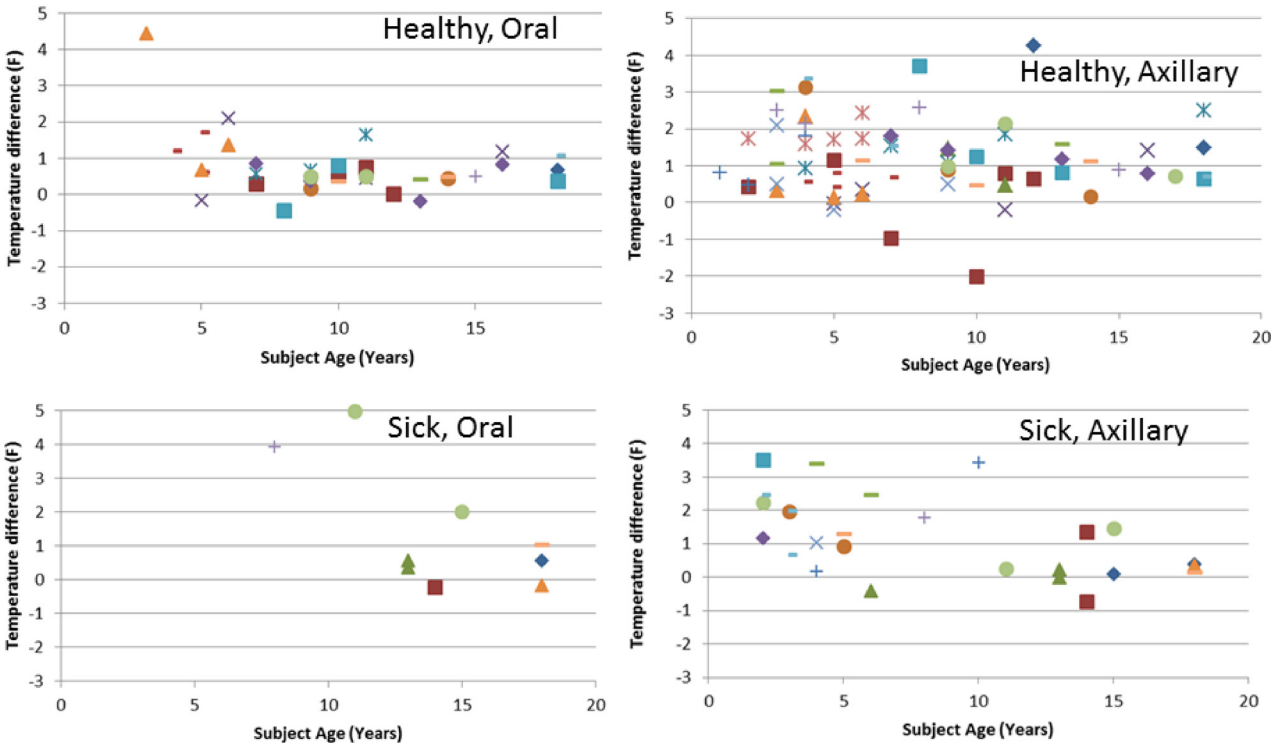


Fig. 12 Continued

References

[1] Crawford, D. C., Hicks, B., and Thompson, M. J., 2006, "Which Thermometer? Factors Influencing Best Choice for Intermittent Clinical Temperature Assessment," J. Med. Eng. Technol., **30**(4), pp. 199–211.

[2] Bock, M., Hohlfeld, U., von Engeln, K., Meier, P. A., Motsch, J., and Tasman, A. J., 2005, "The Accuracy of a New Infrared Ear Thermometer in Patients Undergoing Cardiac Surgery," Can. J. Anesth., **52**(10), pp. 1083–1087.

[3] Cattaneo, C., Frank, S. M., Hesel, T. W., El-Rahmany, H. K., Kim, L. J., and Tran, K. M., 2000, "The Accuracy and Precision of Body Temperature Monitoring Methods During Regional and General Anesthesia," Anesth. Analg., **90**(4), pp. 938–945.

[4] Matsukawa, T., Kashimoto, S., Ozaki, M., Shindo, S., and Kumazawa, T., 1996, "Temperatures Measured by a Deep Body Thermometer (Coretemp®) Compared With Tissue Temperatures Measured at Various Depths Using Needles Placed Into the Sole of the Foot," Eur. J. Anaesthesiol., **13**(4), pp. 340–345.

[5] Erickson, R. S., and Meyer, L. T., 1994, "Accuracy of Infrared Ear Thermometry and Other Temperature Methods in Adults," Am. J. Crit. Care, **3**(1), pp. 40–54.

[6] Hebbar, K., Fortenberry, J. D., Rogers, K., Merritt, R., and Easley, K., 2005, "Comparison of Temporal Artery Thermometer to Standard Temperature Measurements in Pediatric Intensive Care Unit Patients," Pediatr. Crit. Care Med., **6**(5), pp. 557–561.

[7] Hooper, V. D., and Andrews, J. O., 2006, "Accuracy of Noninvasive Core Temperature Measurement in Acutely Ill Adults: The State of the Science," Biol. Res. Nurs., **8**(1), pp. 24–34.

[8] Lawson, L., Bridges, E. J., Ballou, I., Eraker, R., Greco, S., Shively, J., and Sochulak, V., 2007, "Accuracy and Precision of Noninvasive Temperature Measurement in Adult Intensive Care Patients," Am. J. Crit. Care, **16**(5), pp. 485–496.

[9] Smith, L. S., 2003, "Reexamining Age, Race, Site, and Thermometer Type as Variables Affecting Temperature Measurement in Adults—A Comparison Study," BMC Nurs., **2**(1), pp. 1–14.

[10] Weiss, M. E., and Richards, M. T., 1994, "Accuracy of Electronic Axillary Temperature Measurement in Term and Preterm Neonates," Neonat. Network, **13**(8), pp. 35–40.

[11] Jensen, B. N., Jensen, F. S., Madsen, S. N., and Lossl, K., 2000, "Accuracy of Digital Tympanic, Oral, Axillary, and Rectal Thermometers Compared With Standard Rectal Mercury Thermometers," Eur. J. Surg., **166**(11), pp. 848–851.

[12] Kara, A., Devrim, I., Gemngiz, A. B., Celik, F., Tezer, H., Uludag, A. K., and Secmeer, G., 2009, "Is the Axilla the Right Site for Temperature Measurement in Children by Chemical Thermometer?," Turk. J. Pediatr., **51**(4), pp. 325–327.

[13] Dollberg, S., Mincis, L., Mimouni, F. B., Ashbel, G., and Barak, M., 2003, "Evaluation of a New Thermometer for Rapid Axillary Temperature Measurement in Preterm Infants," Am. J. Perinatol., **20**(4), pp. 201–204.

[14] Kitsommart, R., and Phatthanasiriwetin, S., 2005, "Accuracy and Precision of Digital Thermometer in Neonatal Temperature Measurement," Siriraj Med., **57**(5), pp. 128–130.

[15] O'Brien, D. L., Rogers, I. R., Holden, W., Jacobs, I., Mellett, S., Wall, E. J., and Davies, D., 2000, "The Accuracy of Oral Predictive and Infrared Emission Detection Tympanic Thermometers in an Emergency Department Setting," Acad. Emerg. Med., **7**(9), pp. 1061–1064.

[16] Vesnovsky, O., Topoleski, L. D. T., Grossman, L. W., Casamento, J. P., and Zhu, L., 2014, "Performance Testing of Fast Read Digital Thermometers," ASME J. Med. Devices, **8**(3), p. 030905.

[17] Vesnovsky, O., Topoleski, L. D. T., Grossman, L. W., Casamento, J. P., and Zhu, L., 2013, "Evaluation of Temperature Transients at Various Body Temperature Measuring Sites Using a Fast Response Thermistor Bead Sensor," ASME Paper No. SBC2013-14065.

[18] Pompei, F., and Pompei, M., 2004, "Noninvasive Temporal Artery Thermometry: Physics, Physiology, and Clinical Accuracy," Proc. SPIE, **5405**, pp. 61–67.

[19] Suleman, M.-I., Doufas, A. G., Akca, O., Ducharme, M., and Sessler, D. I., 2002, "Insufficiency in a New Temporal-Artery Thermometer for Adult and Pediatric Patients," Anesth. Analg., **95**(1), pp. 67–71.

[20] Latman, N. S., 2003, "Clinical Thermometry: Possible Causes and Potential Solutions to Electronic, Digital Thermometer Inaccuracies," Biomed. Instrum. Technol., **37**(3), pp. 190–196.

[21] Bland, J. M., and Altman, D. G., 1986, "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement," Lancet, **1**(8476), pp. 307–310.

[22] Bland, J. M., and Altman, D. G., 1995, "Comparing Methods of Measurement: Why Plotting Difference Against Standard Method is Misleading," Lancet, **346**(8982), pp. 1085–1087.

[23] Bergman, T., Lavine, A., Incopera, F., and Dewitt, D., 2011, *Fundamentals of Heat and Mass Transfer*, Wiley, Hoboken, NJ.

[24] Saharhkiz, A., Vesnovsky, O., Casamento, J. P., Grossman, L. W., Topoleski, L. D. T., and Zhu, L., 2015, "Evaluating Accuracy of the Algorithm/Formula Used in Store Brand Digital Thermometers to Predict Body Temperature Based on the First Few Seconds of the Temperature Measurements," Summer Biomechanics, Bioengineering, & Biotransport Conference, Snowbird, UT, Paper No. 15-A-507-SB3C.