Journal of Clinical Monitoring and Computing (2019) 33:911–916
https://doi.org/10.1007/s10877-018-0233-1

ORIGINAL RESEARCH

A retrospective evaluation of the risk of bias in perioperative temperature metrics

Robert E. Freundlich[1,2] · Sara E. Nelson[1] · Yuxuan Qiu[3] · Jesse M. Ehrenfeld[1,2,4,5] · Warren S. Sandberg[1,2,4] · Jonathan P. Wanderer[1,2]

Received: 29 October 2018 / Accepted: 5 December 2018 / Published online: 8 December 2018

Abstract
The prevention and treatment of hypothermia is an important part of routine anesthesia care. Avoidance of perioperative hypothermia was introduced as a quality metric in 2010. We sought to assess the integrity of the perioperative hypothermia metric in routine care at a single large center. Perioperative temperatures from all anesthetics of at least 60 min duration between January 2012 and 2017 were eligible for inclusion in analysis. Temperatures were displayed graphically, assessed for normality, and analyzed using paired comparisons. Automatically-recorded temperatures were obtained from several monitoring sites. Provider-entered temperatures were non-normally distributed, exhibiting peaks at temperatures at multiples of 0.5 °C. Automatically-acquired temperatures, on the other hand, were more normally distributed, demonstrating smoother curves without peaks at multiples of 0.5 °C. Automatically-acquired median temperature was highest, 36.8 °C (SD = 0.8 °C), followed by the three manually acquired temperatures (nurse-documented postoperative temperature, 36.5 °C [SD = 0.6 °C]; intraoperative manual temperature, 36.5 °C [SD = 0.6 °C]; provider-documented postoperative temperature, 36.1 °C [SD = 0.6 °C]). Provider-entered temperatures exhibit values that are unlikely to represent a normal probability distribution around a central physiologic value. Manually-entered perioperative temperatures appear to cluster around salient anchoring values, either deliberately, or as an unintended result driven by cognitive bias. Automatically-acquired temperatures may be superior for quality metric purposes.

Keywords Perioperative informatics · Hypothermia · Quality improvement · Electronic health records · Temperature

1 Introduction

Hypothermia, commonly defined as a body temperature below 36 °C, occurs as body heat redistributes from the core to the periphery, caused by vasodilation following the induction of anesthesia [1]. Intraoperative hypothermia is associated with multiple adverse outcomes, including surgical site infection, delayed wound healing, prolonged hospitalization, coagulopathy, and cardiac arrhythmia [2–7]. However, studies have shown that hypothermia during the first hour of anesthesia remains common, even in patients who are actively warmed [8].

To incentivize maintenance of perioperative normothermia, the Physician Quality Reporting System (PQRS) introduced perioperative temperature management as a quality metric in 2010. To meet compliance with this metric, all surgeries longer than 60 min in duration and employing general or neuraxial anesthesia should provide evidence of active intraoperative warming. If active warming is not used, at least one body temperature should measure ≥ 36 °C within a timeframe of 30 min before to 15 min after the documented conclusion of anesthetic care [9]. In 2016, this metric was updated, revising the temperature threshold to ≥ 35.5 °C and eliminating the option for only

✉ Robert E. Freundlich
Robert.e.freundlich@vanderbilt.edu

1 Department of Anesthesiology, Vanderbilt University Medical Center, 1211 21st Avenue South, MAB 422F, Nashville, TN 37212, USA

2 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

3 Department of Emergency Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, USA

4 Department of Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

5 Department of Health Policy, Vanderbilt University Medical Center, Nashville, TN, USA

reporting "usage of active warming" [10]. A performance report is routinely generated by PQRS for participating hospitals on management of perioperative temperatures.

In addition to their role in performance reporting, PQRS quality measures may be used for determining hospital reimbursement. In 2015, the PQRS program introduced a financial penalty for participating practitioners and groups who failed to report data on their quality measures [11]. While the stated goal of this metric was to improve patient safety and provide a standardized outcome measure, each institution is tasked with self-reporting perioperative temperature management to the PQRS. Self-reporting is prone to both deliberate and accidental error and, given the incentives tied to meeting metrics, there is the possibility that clinicians consciously or unconsciously select and record temperatures that meet a performance target. Similar behaviors have been shown in previous studies. One study performed by Wax and colleagues, for example, suggests that manual editing of automatically-recorded data in an Anesthesia Information Management Systems (AIMS) is a common practice [12]. This manual editing may be reinforced by "terminal digit bias", the propensity most clinicians have to bias manually-recorded values towards central anchors such as ".0" and ".5" [13].

To our knowledge, there are currently no data demonstrating accidental or deliberate inaccuracies in the temperatures recorded to satisfy the hypothermia metric in the perioperative literature. However, recognizing the potential for error, we sought to assess whether there is evidence at our institution of strategic or subconscious rounding of perioperative temperature values towards the publicized standard. To do so we examined and compared the distribution of temperature measurements taken preoperatively, intraoperatively, and post operatively. Measurements were taken and recorded both manually and automatically via anesthesia monitors. We hypothesized that the provider-entered perioperative temperature curves would assume a less-normal distribution when compared to the automated temperature curves and would have spikes at increments of 0.5 °C, consistent with terminal digit bias.

2 Methods

The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines were used in the preparation of this manuscript [14].

2.1 Human subjects protection

This study received a quality improvement exemption from the Vanderbilt University Medical Center Institutional Review Board.

2.2 Patient selection

Perioperative data were extracted from the Vanderbilt Perioperative Information Management System database for all surgical patients undergoing procedures with general anesthesia longer than 60 min in duration between January 1, 2012 and January 4, 2017. Neuraxial anesthesia cases were not selected, as core temperature monitoring was not routinely practiced for patients with neuraxial anesthesia. Cases involving cardiopulmonary bypass, cases that had no recorded temperatures, and cases where any of the patient's recorded preoperative, intraoperative (within the last 30 min of the case), or post-operative temperatures fell below 30 °C or above 42 °C were not selected. All temperatures are collected and documented to the nearest tenth of a degree (0.1 °C) in our medical record.

The data were separated into four groups (Table 1), according to perioperative phase and documentation method—automated versus manual entry. Group 1 consisted of preoperative temperatures that were tympanic, oral, axillary or rectal and recorded manually by pre-operative providers. Group 2 temperatures were intraoperative automated temperatures—esophageal, nasal, bladder, or axillary- measured and recorded automatically by the anesthesia monitor. Group 3 measurements were taken manually post-operatively upon patients' arrival to the PACU and were recorded and documented by nursing staff. Group 4 temperatures consisted of PACU temperatures manually documented separately by the in-room anesthesia provider after hand-off to PACU nurses.

Table 1 Groups

Group name	Temperature type	Who recorded this temperature
1	Pre-operative	Provider: anesthesia provider
2	Intra-operative	Automated monitor
3	Post-operative: PACU	Provider: PACU nurse
4	Post-operative: anesthesiology	Provider: anesthesia provider

2.3 Statistical analyses

A histogram of the distribution of body temperatures along with a density curve was created for each of the four groups and descriptive statistics were calculated for each (Fig. 1). Normality for each group was analyzed by employing a Kolmogorov–Smirnov test, a Cramer-von Mises test, and an Anderson–Darling test. Next, an overall comparison of the temperature distributions among the four phases was performed using a Kruskal–Wallis test, followed by Wilcoxon Signed Rank tests to make pairwise comparisons between each phase. Finally, because the Kruskal–Wallis test is not appropriate to test groups with differently shaped distribution curves, a two-sample Kolmogorov–Smirnov test for equality of distribution functions was executed for each bivariate permutation of the perioperative temperatures.

Statistical analyses were performed using SAS version 9.3 (SAS Institute, Cary NC) and figures were created using R version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria).

3 Results

We analyzed 61,926 pre-operative temperatures, 100,327 intraoperative temperatures, 144,863 PACU temperatures, and 137,840 in-room provider post-operative temperatures, from 517,477 unique anesthetics. A total of 267,444 temperatures were excluded based on our prespecified exclusion criteria. In terms of performance 93.56% of cases at our institution met the criteria for appropriate maintenance of normothermia. Table 2 presents the descriptive statistics for each temperature distribution. The automatically-acquired

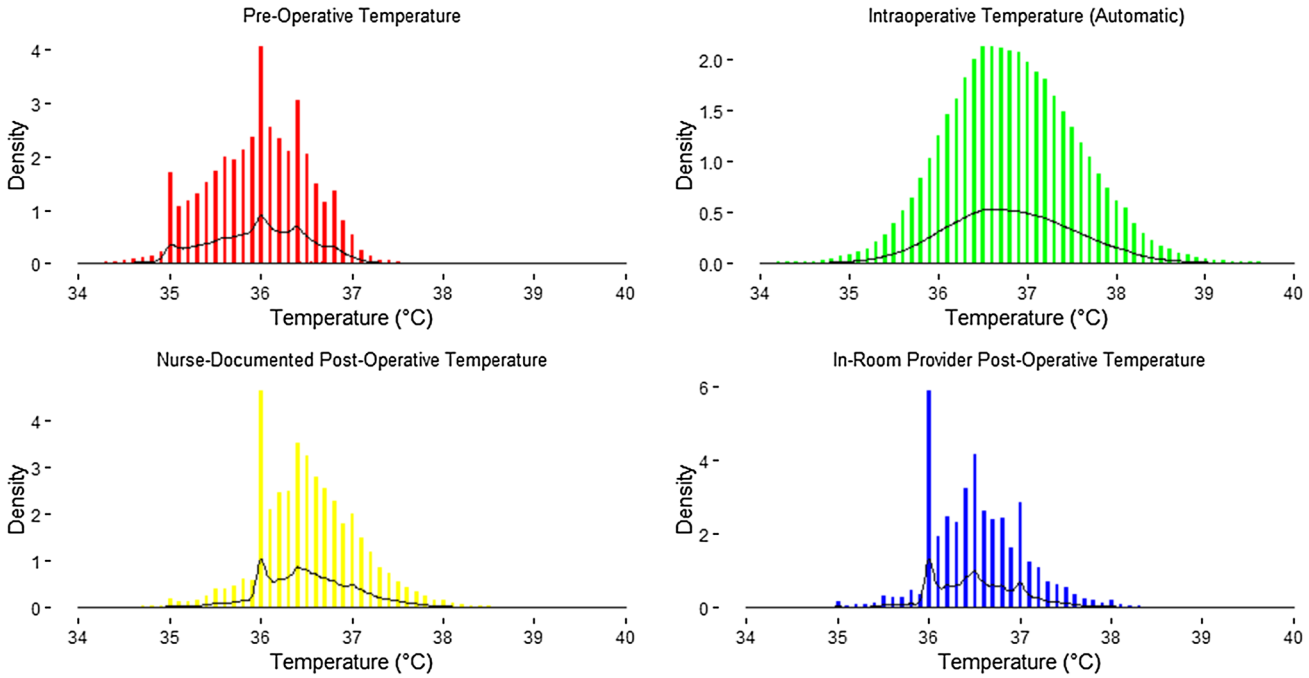

Fig. 1 Histogram of the distribution of temperatures in the four sub-groups

Table 2 Descriptive statistics for each group of perioperative temperatures

	Preoperative temperature	Automated intraoperative temperature	Nurse post-anesthesia care unit temperature	Provider post-anesthesia care unit temperature
Number (n)	61,926	100,327	144,863	137,840
Mean (95% CI)	36.07 (36.07, 36.08)	36.82 (36.82, 36.83)	36.53 (36.53, 36.53)	36.48 (36.48, 36.48)
Median	36.1	36.8	36.5	36.5
Mode	36.0	36.6	36.0	36.0
Standard deviation	0.57	0.78	0.57	0.62

intraoperative median temperature was highest, at 36.80 °C (SD = 0.778 °C), followed by the nurse-documented postoperative median temperature, at 36.50 °C (SD = 0.572 °C), then the anesthesia provider-documented postoperative median temperature at 36.50 °C, (SD = 0.618 °C) and finally, the preoperative median temperature, at 36.10 °C (SD = 0.573 °C). The mode values for the manually-entered temperatures (Groups 1, 3 and 4) were the same at 36.0 °C, while the mode for the machine-entered temperature (Group 2) was 36.6 °C.

Histograms with density curves for temperature distribution of each group are presented side by side for comparison in Fig. 1. Histograms displaying the automated-entry intraoperative temperatures reveals a smoother distribution and density curve when compared to each provider-entered temperature distribution, all of which display pronounced spikes at increments of 0.5 °C.

Normality assessments for the automated intraoperative temperatures and the PACU provider-entry temperatures are reported in Table 3. The Kolmogorov–Smirnov test, Cramer-von Mises test and Anderson–Darling test all produced lower test statistics for the automated-entry group when compared to the manually entered groups, indicating the automated temperature distribution is closer to a normal distribution when compared to provider-entered temperatures.

The Kruskal–Wallis overall comparison produced a p-value of < 0.0001 (Table 1), indicating that at least one group's temperature distribution significantly differs from the others. The pairwise comparisons using Wilcoxon Signed Rank tests, showed significant differences in each pairing ($p < 0.001$ for all) except for the comparison between the postoperative temperatures documented by anesthesia and nursing providers ($p = 0.476$). The above results were further reinforced by the two sample Kolmogorov–Smirnov tests for equality of distribution functions (Table 3), which again found significant differences in each pairing ($p < 0.001$ for all) except for the comparison between the postoperative temperatures documented by anesthesia and nursing providers ($p = 0.31$).

4 Discussion

We present data demonstrating that manually-entered intraoperative temperature distributions differed from the distribution automatically-documented intraoperative temperatures. Automatically-documented temperature distributions were smoother, more normally distributed, and did not exhibit spikes at multiples of 0.5 °C. In other words, putting a human documenter between an electronic instrument and the record introduces error.

4.1 Weaknesses of manually entered data in electronic medical records

To seek an explanation for the significant difference in normality between automatically recorded intraoperative temperatures and provider-entered temperatures, we reviewed the literature on manual entry versus automated entry of anesthesia information. One study of anesthesia start and end times found that paper anesthesia records written by anesthesia providers had significantly more times ending with the digits "0" or "5" than those recorded electronically in an AIMS, an example of so-called "terminal digit bias" [12, 13, 15]. Another similar cross-sectional study showed that automatic time capture in an electronic anesthesia record system had a more uniform distribution of billing times compared to manual entries [15, 16]. Consistent with these findings, our data display a non-smooth distribution in manually recorded data, without much variation, other than a pronounced spike at multiples of 0.5 °C.

One obvious concern about manually-entered data of any kind is that they may be manipulated, either consciously or subconsciously to meet targets. While performance metrics, such as PQRS's Perioperative Temperature Management exist to improve transparency and quality of healthcare delivery, the nature of their use creates incentives for conscious or unconscious data distortion. A prominent example of this is the cost-quality trade-off scandal at the Veterans Administration hospitals, in which artificial reported wait-time distortion was implicated in dozens of deaths [17]. Other instances of biasing performance data have included reclassification of cardiac patient diagnoses and upcoding—coding to make patients appear sicker than they actually are [18]. In surgery, public reporting of cardiac surgeon performance has led to

Table 3 Normality Assessments of Intraoperative versus PACU group

	Preoperative	Intraoperative	PACU Nurse	In-room Provider
Kolmogorov–Smirnov	0.07	0.04	0.09	0.10
Cramer-von Mises	29.88	25.06	123.91	152.13
Anderson–Darling	182.69	153.82	771.14	1037.04
Kurtosis	1.14	1.87	1.46	2.77

the unintended adverse effect of surgeons manipulating their score by refusing to take the necessary cases of high-risk cardiology patients [19, 20].

Artificial distortion of electronic anesthesia record systems has also been documented. One study reported significant information loss in manual compared to automatic-vital sign collection from the same anesthetics, giving the AIMS data a non-smoothed appearance [21]. Another study demonstrated that 7% of systolic blood pressure recordings that were invalidated for being out of range were replaced by provider-entered values [22].

While we did not identify any literature within or outside of the field of medicine that specifically discusses temperature smoothing with the purpose of "gaming", increased evidence of artificial distortion of other performance metrics calls for the examination of perioperative temperature as a candidate for this practice as well. Therefore, when considered alongside these studies, we believe the results of our analysis at least suggest the potential for providers to exhibit either unconscious mental rounding of temperature measurements. Given that manually-documented temperatures were consistently lower, however, we would contend that this is more likely due to factors such as terminal digit bias, rather than a desire to artificially "game" quality metrics.

4.2 Manual vs automated-entry data distributions in this study

The mode of each manual-entry temperature distribution was 36.0 °C, which is 0.6 °C lower than the mode for the automated-entry temperature distribution. We note that the central tendency measures (Mean, Median, Mode) of the manually entered temperature data are all lower (and farther away from the 'normal' value of 37 °C) than their electronically gathered counterparts. This does not seem to support the notion that providers were either consciously or subconsciously targeting a set value in pursuit of a metric. When pairwise comparisons were made, the distribution of recorded temperatures were all significantly different, with the exception Group 3 versus Group 4. Again, this was expected, as the two recordings represent the same measurement logged by two different individuals—the PACU nurse and the in-room anesthesia provider.

Our analysis also showed that automated intraoperative temperatures were significantly higher than PACU and preoperative temperatures. As temperatures normally differ between measuring sites, this may be partially explained by a high frequency of core temperature monitoring in the operating room, which should be expected to be higher than peripheral temperatures, such as skin temperature [23]. Alternatively, this may be evidence that providers were not consciously or subconsciously biasing the data to achieve compliance with quality metrics.

4.3 Strengths and limitations

This study was conducted over 5 years, at a large tertiary care center, and represents a wide variety of providers and surgeries, significantly improving its generalizability to other centers. We introduce a method for evaluating artificial metric distortion that is flexible and easily adapted elsewhere.

A limitation of this study is that we did not assess each provider's awareness of the PQRS Perioperative Temperature Management metric; therefore, we cannot and should not conclude that awareness of the metric engendered artificial distortion of recorded data. In 2016 the PQRS metric goal decreased from 36.0 to 35.5 °C. As there were visible spikes at both values in manually-recorded values, we do not believe that this significantly impacted our findings. As we have discussed elsewhere, temperature in the normal human body is non-uniform, with normal expected variation reflecting a cooler periphery and warmer core. Our AIMS did not document the source of a perioperative temperature and, as such, some of our observed variation may be due to differences in monitoring sites and differences in accuracy between different types of thermometers. While this may explain differences between automated (often core) and manually-documented (often peripheral) temperatures, it does not explain the markedly-different distributions of these values. Similarly, while patient factors, such as age, may explain variation in temperature, it is unlikely that this would independently introduce bias into documentation patterns.

5 Conclusion

Our analysis demonstrated significantly different temperature distributions between automated-entry temperatures and provider-entered temperatures. There were notably spikes in increments of 0.5 °C for all provider-entry temperatures, but not for automated-entry, suggesting intentional or unintentional rounding of perioperative temperatures. Future work should include qualitative studies focused on anesthesia providers' awareness and understanding of perioperative temperature metrics, as well as behavioral analyses examining provider entry of perioperative temperature.

Acknowledgements Patrick Jablonski, Ph.D., provided statistical recommendations to assist in the preparation and refinement of the statistical analysis.

Funding Dr. Freundlich receives Grant support from an NIH-NCATS KL2 (KL2 TR002245).

Compliance with ethical standards

Conflict of interest Dr. Freundlich has received grant support and consulting fees from Medtronic for work unrelated to the content of this manuscript.

References

1. Hart SR, et al. Unintended perioperative hypothermia. Ochsner J. 2011;11(3):259–70.
2. Madrid E, et al. Active body surface warming systems for preventing complications caused by inadvertent perioperative hypothermia in adults. Cochrane Database Syst Rev. 2016;4:Cd009016.
3. Kurz A, Sessler DI, Lenhardt R. Perioperative normothermia to reduce the incidence of surgical-wound infection and shorten hospitalization. Study of Wound Infection and Temperature Group. N Engl J Med. 1996;334(19):1209–15.
4. Poveda VB, Nascimento AS. The effect of intraoperative hypothermia upon blood transfusion needs and length of stay among gastrointestinal system cancer surgery. Eur J Cancer Care (Engl), 2017. 26(6).
5. Yi J, et al. Maintaining intraoperative normothermia reduces blood loss in patients undergoing major operations: a pilot randomized controlled clinical trial. BMC Anesthesiol. 2018;18(1):126.
6. Wong PF, et al. Randomized clinical trial of perioperative systemic warming in major elective abdominal surgery. Br J Surg. 2007;94(4):421–6.
7. Alderson P, et al. Thermal insulation for preventing inadvertent perioperative hypothermia. Cochrane Database Syst Rev, 2014(6): Cd009908.
8. Sun Z, et al. Intraoperative core temperature patterns, transfusion requirement, and hospital duration in patients warmed with forced air. J Am Soc Anesthesiol. 2015;122(2):276–85.
9. PQRS Measure #193: Perioperative Temperature Management. 2015. https://www.pqrspro.com/cmsmeasures/2015/perioperative_temperature_management/. Accessed 25 Mar 2017.
10. PQRS Measure #424: Perioperative Temperature Management. 2016. https://www.pqrspro.com/cmsmeasures/2016/perioperative_temperature_management-2/. Accessed 25 Mar 2017.
11. Medicare CF, Services M. Physician quality reporting system. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html. Accessed 21 June 2012.
12. Wax DB, et al. Manual editing of automatically recorded data in an anesthesia information management system. J Am Soc Anesthesiol. 2008;109(5):811–5.
13. Thavarajah S, White WB, Mansoor GA. Terminal digit bias in a specialty hypertension faculty practice. J Hum Hypertens. 2003;17(12):819–22.
14. von Elm E, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–7.
15. Deal LG, et al. Are anesthesia start and end times randomly distributed? The influence of electronic records. J Clin Anesth. 2014;26(4):264–70.
16. Phelps M, et al. Comparison of minute distribution frequency for anesthesia start and end times from an anesthesia information management system and paper records. J Clin Monit Comput. 2016: 31: 1–6.
17. Slutsky AS, Ranieri VM. Mechanical ventilation: lessons from the ARDSNet trial. Respir Res. 2000;1(2):73–7.
18. Chatterjee P, Joynt KE. Do cardiology quality measures actually improve patient outcomes? J Am Heart Assoc. 2014;3(1):e000404.
19. Burack JH, et al. Public reporting of surgical mortality: a survey of New York State cardiothoracic surgeons. Ann Thor Surg. 1999;68(4):1195–200.
20. Hua M, et al. Impact of public reporting of 30-day mortality on timing of death after coronary artery bypass graft surgery. Anesthesiology. 2017;127(6):953–60.
21. Van Schalkwyk J, et al. Does manual anaesthetic record capture remove clinically important data? Br J Anaesthesia. 2011;107(4):546–52.
22. Block FE. Normal fluctuation of physiologic cardiovascular variables during anesthesia and the phenomenon of "smoothing". J Clin Monit. 1991;7(2):141–5.
23. Hermanns H, et al. Assessment of skin temperature during regional anaesthesia-What the anaesthesiologist should know. Acta Anaesthesiol Scand. 2018;62(9):1280–9.