International Journal of Biometeorology (2025) 69:2405–2409
https://doi.org/10.1007/s00484-025-02957-w

SHORT COMMUNICATION

Performance of summary measures of core body temperature in heat-related health outcome regression models

M. Dally[1,2] · A. Dye-Robinson[1,3] · Y. Li[1,2] · D. Jaramillo[1] · K. Villarreal-Hernandez[1] · L. Krisher[1,2] · J. Butler-Dawson[1,2] · J. Schaeffer[4] · J. L. Adgate[1,2] · K. A. James[1,2,5] · L. S. Newman[1,2,5,6]

Received: 9 July 2024 / Revised: 22 April 2025 / Accepted: 25 May 2025 / Published online: 12 June 2025

Abstract
Climate-related hazards are a leading health and safety concern for working populations. Technological advancements have made real-time collection of core body temperature, a measure of heat strain, more feasible in non-clinical settings. Researchers and practitioners are now faced with how to best summarize a plethora of core body temperature data. Core body temperature data were collected over the course of the workday for 128 agricultural workers in Guatemala from December 2021 to April 2023. Concurrently, measures of heat-related health effects (reduced kidney function, muscle breakdown, and dehydration) were collected. Using linear mixed-effect regression models we assessed the performance of 14 summary measures of core body temperature on the heat-related health effects of interest. Model performance was compared using AIC and BIC. There was no consistent summary statistic that provided the best fit based on AIC or BIC for all four heat-related health effects. The difference between the observed CBT and the average of the first 60 min of the cleaned workday data (delta) provided the best performance for percent change in creatine kinase (muscle breakdown), serum creatinine (reduced kidney function), and serum osmolality (dehydration). The area under the curve (AUC) performed best for dichotomous indicator of acute kidney injury. Choice of how to operationalize longitudinal measurements of core body temperature generated using ingestible core body temperature pills should be driven by the research question and health effect of interest.

Keywords Occupational health · Heat stress · Workplace · Agricultural workers

✉ M. Dally
miranda.dally@cuanschutz.edu

1 Centers for Health, Work & Environment, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

2 Department of Environmental and Occupational Health, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

3 Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

4 Department of Environment and Radiological Health Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA

5 Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

6 Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Introduction

The impacts of climate-related hazards on human health are at the forefront of public health, necessitating improvements in measurement of the body's response to environmental exposures, especially heat. To understand the mechanistic role and attributable risk of heat exposure on morbidity and mortality (Ebi et al. 2021), we must assess an individual's response to heat stress, referred to as heat strain (Lemke and Kjellstrom 2012). Heat strain can be measured by a number of physiological symptoms, including core body temperature (CBT) (Sawka et al. 1992). The human body endeavors to maintain a CBT of around 37 °C by modulating heat generation and heat dissipation (Dally and Newman 2021).

Until recently, measuring the CBT outside of a clinical setting has been difficult (Hymczak et al. 2021). One technological advance has been the development of the so-called CBT pill, which can be used in non-clinical settings. As they pass through the gastrointestinal tract, these telemetric ingestible capsules collect and transmit internal temperature data that is downloadable (Pearson et al. 2022). Faced with a surfeit of longitudinal CBT data, researchers must decide how to best utilize the information in epidemiologic investigations of health effects.

Researchers and public health practitioners have taken several approaches to summarizing longitudinal measures into a single summary statistic, often at a threshold of 38.0 °C (NIOSH 2016). However, there remains uncertainty concerning which estimate of CBT (e.g., peak, average, cumulative temperature) best predicts downstream health effects. To address this need, we conducted empiric research in the field setting and compared how well various summary measures of CBT perform in regression models that assess clinical indicators that have been associated with heat strain, specifically dehydration, muscle injury, and kidney function (Fortes et al. 2013; Moyce et al. 2017).

Methods

For this analysis we utilized data previously collected as part of a larger study examining the association of particulate air pollution and heat on kidney function of male sugarcane cutters in Southwest Guatemala (the Chronic Kidney Disease, Heat and Air Pollution Study, CHAP I)(grant R01 ES 031585). In brief, during six time points spanning from early harvest in December 2021 to late harvest in April 2023, we performed venipuncture to obtain serum samples from workers both pre- and post-work shift, while concurrently measuring workers' CBT (Bodycap e-Celsius Performance®, Caen, France). Each participating worker could have up to 3 days of data collected. Ethical approval for this study was provided by the Colorado Multiple Institutional Review Board (#20–0509) and the Comité de Ética Independiente ZUGUEME in Guatemala.

Utilizing a wireless, external receiver (e-Viewer, BodyCap, Caen, France), CBT temperature data were collected continuously from the pill in 2-minute intervals during the approximate 9–10-hour workday. Data were then cleaned to exclude the first 60 min of measurements to account for the time between pill ingestion and transit to the lower gastrointestinal tract and for the readings to stabilize (Moyce et al. 2017; Pearson et al. 2022). Additionally, following the methods of Petropoulos et al. 2023 we removed individual CBT readings where the average slope between a reading and its neighboring readings was greater than two standard deviations away from the mean of all slopes for the worker that day and the value of the individual CBT reading was more than two standard deviations away from the mean of all CBT values for the worker on that day (Petropoulos et al. 2023). CBT readings were also removed when the absolute slope between a given reading and neighboring readings was greater than the equivalent of a 2 °C change over 15 min. Further, individuals who were missing >30% of CBT readings for a given day were excluded from the analysis.

Health outcomes

The health outcomes of interest for this analysis were markers of decline in kidney function, hydration status, and muscle breakdown. Serum creatinine (Cr) was measured using the i-STAT handheld (Abbott, Princeton, NJ, USA). Decline in kidney function was examined as (1) the continuous measure of percent change in Cr from pre- to post-shift and (2) the dichotomous indicator of acute kidney injury, defined as a greater than 0.3 mg/dL change in Cr from pre- to post-shift (Goyal et al. 2024). Hydration status was assessed as a continuous variable defined by percent rise in serum osmolality (Osmo) from pre- to post-shift. Muscle breakdown was defined as the percent change in serum creatine kinase (CK) from pre- to post-shift.

Summary measures of CBT

We compared 14 summary measures of CBT during the workday. While this is not an exhaustive list of potential ways to summarize CBT this list is reflective of some of the common measurements found in the literature (McKenna et al. 2023; Moyce et al. 2020; Petropoulos et al. 2023; Pryor et al. 2020). These included: (1) the average of all data points collected in 2-minute intervals (2 min average); (2) the maximum of all data points collected in 2-minute intervals (2 min maximum); (3) the average of data points collected in 4-minute intervals (two 2-minute intervals combined; 4 min average); (4) the maximum of all data points collected in 4-minute intervals (4 min maximum); (5) the difference between the 2 min maximum and the average of the first 60 min of the cleaned work day data (delta); (6) the area under the curve covering the entire workday calculated using a trapezoid approximation (overall AUC); (7) the area under the curve for measurements above the 75 th percentile of a worker's measurements (AUC over 75 th percentile); (8) the area under the curve for measurements above the 85 th percentile of a worker's measurements (AUC over 85 th percentile); (9) the area under the curve for measurements above the 95 th percentile of a worker's measurements (AUC over 95 th percentile); (10) whether a worker's CBT exceeded 37 °C at any point during the workday (above 37); (11) whether a worker's CBT exceeded 37.5 °C at any point during the workday (above 37.5); (12) whether a worker's CBT exceeded 38 °C at any point during the workday (above 38); (13) whether a worker's CBT exceeded 38.5 °C at any point during the workday (above 38.5); and (14) whether a worker's CBT exceeded 39 °C at any point during the workday (above 39).

Statistical analysis

To assess the performance of each of the 14 summary statistics, we ran univariate mixed-effects simple linear regression models for continuous health outcomes and univariate mixed-effects logistic regression models for dichotomous outcomes. Each mixed-effect model included a random intercept for each worker. For the 56 models, we compared the resulting Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Dziak et al. 2020). Models with the lowest AIC and BIC were considered the best performing models. All analyses were conducted in R version 4.3.1 (Vienna, Austria) (R Core Team 2017).

Results

There were 195 observations from 128 workers available for this analysis. The average mean environmental temperature during the study period was 30 °C (SD: 4). The average mean percent change in Cr was 12% (SD: 15%), in CK was 42% (SD: 31%), and in Osmo was −2% (SD: 3%). There were 20 (9%) observed pre- to post-shift AKI.

Table 1 Fit statistic comparison of the mixed-effects regression models for the 14 summary statistics with the four health outcomes of interest. Presented as AIC/BIC

	Health Outcomes of Interest			
Summary Statistic	Percent Change in Serum Creatinine	Acute Kidney Injury	Percent Change in Creatine Kinase	Percent Change in Serum Osmolality
2 min Average	1500/1512	125/135	946/958	1716/1728
2 min Maximum	1500/1513	125/134	946/959	1716/1729
4 min Average	1499/1512	125/135	946/958	1716/1728
4 min Maximum	1500/1513	125/134	946/959	1717/1729
Delta	**1494/1507**	124/133	**941/953**	**1708/1721**
Overall AUC	1505/1517	127/136	947/960	1718/1731
AUC Over 75 th Quantile	1502/1515	126/136	947/960	1718/1731
AUC Over 85 th Quantile	1503/1516	126/135	947/960	1717/1730
AUC Over 95 th Quantile	1503/1516	**120/130**	947/959	1719/1731
Above 37	1505/1518	*	943/956	1718/1731
Above 37.5	1498/1511	127/136	945/958	1719/1731
Above 38	1499/1512	121/131	946/959	1718/1731
Above 38.5	1502/1515	126/135	942/955	1715/1728
Above 39	1503/1515	126/136	947/959	1718/1731

*Indicates the model did not converge

Bolded numbers represent the lowest AIC/BIC values in each column

Table 1 provides the AIC and BIC for each of the summary statistics for each of the health outcomes. There was no consistent summary statistic that provided the best fit based on AIC or BIC. The AUC over 95 th percentile provided the best performance for a dichotomous indicator of AKI. For percent change in Cr, percent change in CK, and percent change in Osmo the delta statistic provided the best fit.

Looking specifically at the measurements of body temperature threshold (Fig. 1), indicators for CBT exceeding 37.5 °C provided the best fit for percent change in Cr, 38 °C provided the best fit for AKI, and 38.5 °C provided the best fit for percent change in CK and percent change in Osmo.

Discussion

No single summary measure of daily changes in CBT provided the best fit when examining four different heat-associated health effects in a cohort of agricultural workers performing high exertion work under hot environmental conditions. Choice of how to operationalize longitudinal measurements of CBT generated using ingestible CBT pills should be driven by the research question and health effect of interest.

Our analysis suggests that change from start of work CBT to maximum CBT over the course of the work-shift (Delta) may be the most appropriate statistic for epidemiological studies examining the role of elevated CBT on health outcomes. The Delta statistic performed the best for percent change in Cr, percent change in CK, and percent change in Osmo while the AUC statistic performed the best for AKI. Both Delta and AUC are an overall measure of heat strain throughout the day. When used in statistical models, the interpretation of Delta and AUC can address the overall burden of heat on the selected outcome. Additional methods that leverage the full information, such as spline regression, warrant further consideration.

Threshold statistics are commonly used in heat mitigation policies and harm reduction programs (NIOSH 2016). Unlike Delta and AUC, thresholds are more easily monitored in real time, allowing for immediate risk reduction interventions. In this analysis, we see that the performance of various thresholds was inconsistent among health outcomes, which suggests the need to a priori determine which metric provides the best fit for the targeted health outcome.

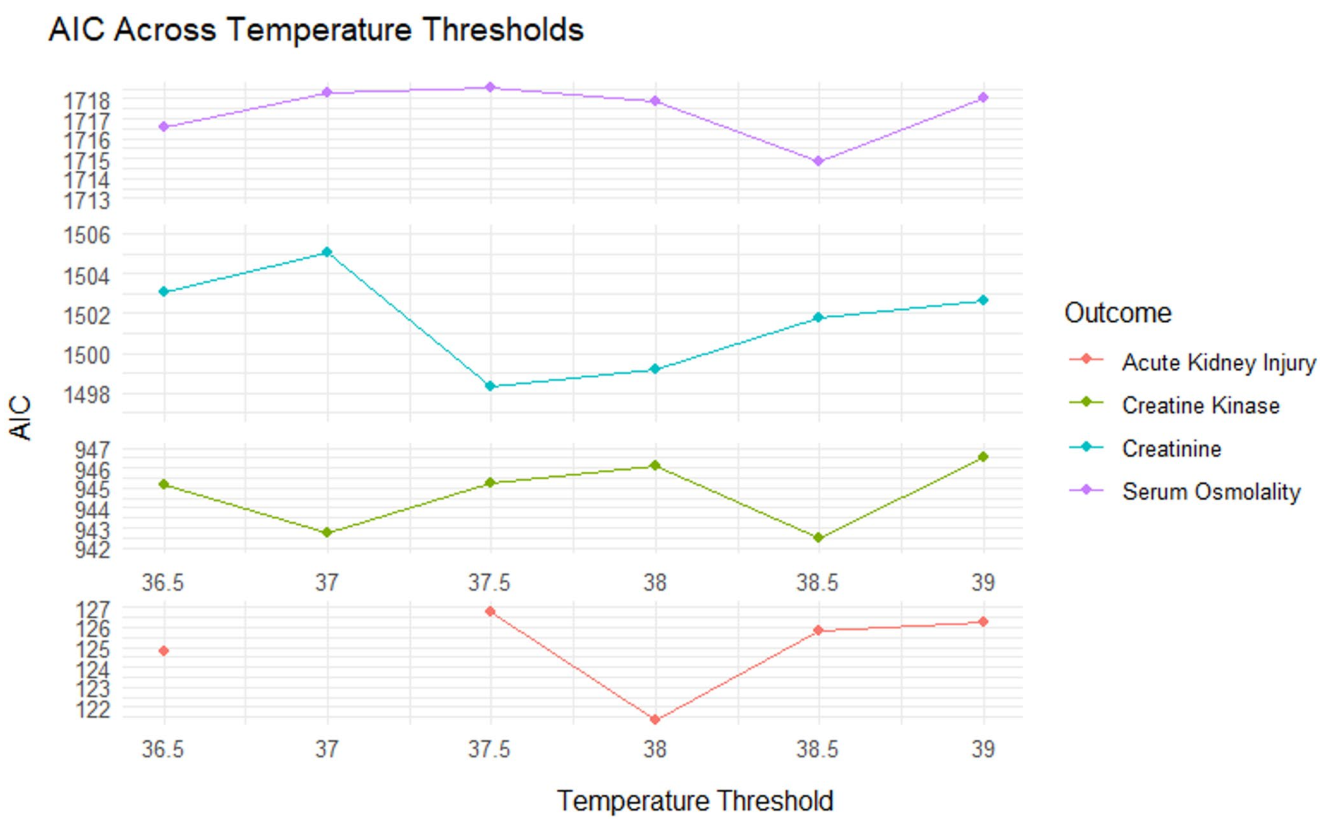

Fig. 1 AIC performance of CBT thresholds on the four health outcomes of interest from mixed-effects regression models

Limitations

The major limitation of our study is that we only examined four health outcomes. While these four health outcomes are commonly associated with heat stress (Fortes et al. 2013; Moyce et al. 2017), they are not an exhaustive list. Secondly, there are additional ways to summarize CBT that we did not explore. Thirdly, our study was conducted only among men who sustained heavy work over the course of multiple hours. Additional research is needed to examine the role of sex, other demographic factors, and workload on elevations of CBT and associated health outcomes. Although AIC and BIC are widely used for model selection in public health research (Dziak et al. 2020), they were the only fit statistics we examined.

Conclusions

This study suggests that the optimal summary measure of core body temperature should be determined based on the health outcome of interest. When examining the association between core body temperature and health outcomes there is a need to account for other predictive factors, such as smoking status, job tasks, and demographics (Butler-Dawson et al. 2018), that could influence the health outcome of interest.

Author contributions MD– Conceptualization, supervision, methodology, writing original draft; ADR– Formal analysis, software, data curation, writing original draft; YL– data curation, writing review & editing; DJ– project administration, investigation, writing review & editing; KVH– project administration, investigation, writing review & editing; LK– project administration, investigation, writing review & editing; JBD– project administration, investigation, writing review & editing; JS– resources, funding acquisition, writing review & editing; JLA– resources, funding acquisition, writing review & editing; KAJ– methodology, writing review & editing; LSN– conceptualization, supervision, resources, funding acquisition, writing review & editing.

Funding This study was supported by National Institutes of Health (NIH) (R01 ES31585) and Training Grant T42 OH00929 funded by the Centers for Disease Control and Prevention. Funders had no role in data analysis, interpretation of data, writing the manuscript, or the decision to submit the findings for publication.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The University of Colorado has a memorandum of agreement with Pantaleon, a Guatemala-based agribusiness. Pantaleon has provided access to the employees who volunteered to participate in this research project. The University of Colorado employed appropriate research methods in keeping with academic freedom, based conclusions on critical analysis of the evidence and reported findings fully and objectively. The terms of this arrangement have been reviewed and approved by the University of Colorado in accordance with its conflict of interest policies.

References

Butler-Dawson J, Krisher L, Asensio C, Cruz A, Tenney L, Weitzenkamp D, Dally M, Asturias EJ, Newman LS (2018) Risk factors for declines in kidney function in sugarcane workers in Guatemala. J Occup Environ Med 60(6):548–558. https://doi.org/10.1097/JOM.0000000000001284

Dally M, Newman LS (2021) Worker health. In: Lemery J, Knowlton K, Sorenson CJ (eds) Global climate change and human health, 2nd edn. Wiley

Dziak JJ, Coffman DL, Lanza ST, Li R, Jermiin LS (2020) Sensitivity and specificity of information criteria. Brief Bioinform 21(2):553–565. https://doi.org/10.1093/bib/bbz016

Ebi KL, Capon A, Berry P, Broderick C, de Dear R, Havenith G, Honda Y, Kovats RS, Ma W, Malik A, Morris NB, Nybo L, Seneviratne SI, Vanos J, Jay O (2021) Hot weather and heat extremes: health risks. Lancet 398(10301):698–708. https://doi.org/10.1016/s0140-6736(21)01208-3

Fortes MB, Di Felice U, Dolci A, Junglee NA, Crockford MJ, West L, Hillier-Smith R, Macdonald JH, Walsh NP (2013) Muscle-damaging exercise increases heat strain during subsequent exercise heat stress. Med Sci Sports Exerc 45(10):1915–1924. https://doi.org/10.1249/MSS.0b013e318294b0f8

Goyal A, Daneshpajouhnejad P, Hashmi M, Bashir K (2024) Acute kidney injury. In: StatPearls [Internet]. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK441896/

Hymczak H, Golab A, Mendrala K, Plicner D, Darocha T, Podsiadlo P, Hudziak D, Gocol R, Kosinski S (2021) Core temperature measurement-principles of correct measurement, problems, and complications. Int J Environ Res Public Health 18(20). https://doi.org/10.3390/ijerph182010606

Lemke B, Kjellstrom T (2012) Calculating workplace WBGT from meteorological data: a tool for climate change assessment [Review]. Ind Health 50(4):267–278. https://doi.org/10.2486/indhealth.MS1352

McKenna ZJ, Nava RC, Magalhaes C, Magalhaes de Castro F, Machado Moreira C, Shibuya E, da Cunha I, Maia PA, Cauduro Roscani R, Rocha-Vieira E, Santiago Moraes Barros TM, Bitencourt D, Amorim FT (2023) A comprehensive evaluation of heat stress and heat strain in a sample of sugarcane cutters in Brazil. J Appl Physiol (1985) 135(6):1431–1439. https://doi.org/10.1152/japplphysiol.00555.2023

Moyce S, Mitchell D, Armitage T, Tancredi D, Joseph J, Schenker M (2017) Heat strain, volume depletion and kidney function in California agricultural workers. Occup Environ Med 74(6):402–409. https://doi.org/10.1136/oemed-2016-103848

Moyce S, Armitage T, Mitchell D, Schenker M (2020) Acute kidney injury and workload in a sample of California agricultural workers. Am J Ind Med 63(3):258–268. https://doi.org/10.1002/ajim.23076

NIOSH (2016) NIOSH criteria for a recommended standard: occupational exposure to heat and hot environments

Pearson SJ, Highlands B, Jones R, Matthews MJ (2022) Comparisons of core temperature between a telemetric pill and heart rate estimated core temperature in firefighters. Saf Health Work 13(1):99–103. https://doi.org/10.1016/j.shaw.2021.11.003

Petropoulos ZE, Keogh SA, Jarquin E, Lopez-Pilarte D, Amador Velazquez JJ, Garcia-Trabanino R, Sanchez A, Guevara MR, Gruener R, Allen A, Leibler DR, Delgado JH, McClean IS, Friedman MD, Brooks DJ, Scammell MK (2023) Heat stress and heat strain among outdoor workers in El Salvador and Nicaragua. J Expo Sci Environ Epidemiol 33(4):622–630. https://doi.org/10.1038/s41370-023-00537-x

Pryor RR, Pryor JL, Vandermark LW, Adams EL, Brodeur RM, Schlader ZJ, Armstrong LE, Lee EC, Maresh CM, Casa DJ (2020) Acute kidney injury biomarker responses to short-term heat acclimation. Int J Environ Res Public Health 17(4). https://doi.org/10.3390/ijerph17041325

R Core Team (2017) R: a language and environment for statistical computing. In: R Foundation for Statistical Computing. https://www.R-project.org/

Sawka MN, Young AJ, Latzka WA, Neufer PD, Quigley MD, Pandolf KB (1992) 1992/07/01). Human tolerance to heat strain during exercise: influence of hydration. J Appl Physiol 73(1):368–375. https://doi.org/10.1152/jappl.1992.73.1.368